AD_____

Award Number: DAMD17-96-1-6172

TITLE: Mechanism of Splicing of Unusual Intron in Human Proliferating Cell Nucleolar P120

PRINCIPAL INVESTIGATOR: Michelle Hastings, Ph.D.
Adrian Krainer, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

REPORT DATE: December 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1999	Final (1 Dec 96 - 30 Nov 99)

4. TITLE AND SUBTITLE
Mechanism of Splicing of Unusual Intron in Human Proliferating Cell Nucleolar P120

5. FUNDING NUMBERS
DAMD17-96-1-6172

6. AUTHOR(S)
Michelle Hastings, Ph.D.
Adrian Krainer, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

E-MAIL:
krainer@cshl.org
hastings@cshl.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The purpose of this project has been to gain a better understanding of pre-mRNA splicing mechanisms by studying SR proteins (a family of trans-acting splicing factors), exonic splicing enhancers and non-conventional splicing of AT-AC introns. AT-AC introns are a rare type of intron present in a wide variety of genes including P120. High P120 gene expression correlates with rapid cell proliferation as in cancer cells. The goals of the proposal have been met by work in three major areas of investigation.

Specific consensus exonic splicing enhancers (ESEs) for individual SR proteins, SF2/ASF, SRp40, SRp55 and SC35, were identified by an in vitro SELEX procedure. Sequences matching these consensus ESEs stimulate splicing of conventional and non-conventional introns.

A nonsense mutation in BRCA1, the breast cancer susceptibility gene, that is associated with familial breast and ovarian cancer causes skipping of the entire exon. This mutation is within a consensus SR protein binding site and *in vitro* splicing analysis showed that exon skipping is due to the failure of an SR protein, SF2/ASF, to recognize the ESE.

Using in vitro splicing complementation assays, SR proteins, which are required for conventional intron splicing, were shown to be required for excision of AT-AC introns.

14. SUBJECT TERMS
Breast Cancer, pre-mRNA splicing, exonic splicing enhancer, BRCA1, AT-AC introns

15. NUMBER OF PAGES
58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

MH Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

X In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

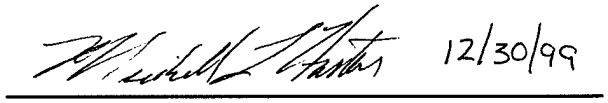


TABLE OF CONTENTS

(5) INTRODUCTION

This research project has investigated general mechanisms of pre-mRNA splicing by studying SR proteins (trans-acting splicing factors), exonic splicing enhancers (ESEs) and non-conventional splicing of AT-AC introns. Recent reports show that alterations in the levels of specific SR proteins have recently been associated with changes in alternative splicing in mammary tumorogenesis and colon adenocarcinomas(Stickeler et al., 1999; Ghigna et al., 1998). The role of individual SR proteins in splicing is not well understood and remains an important question for understanding normal and disease-state pre-mRNA splicing. AT-AC introns are a rare type of intron present in a wide variety of genes including P120. High P120 gene expression correlates with rapid cell proliferation characteristic of cancer cells. The goals of the proposal have been met by work in three major areas of investigation.

The initial recipient and investigator for this project, Dr. Liu, used an in vitro splicing SELEX procedure to identify specific consensus exonic splicing enhancer (ESE)sequences for individual SR proteins, SF2/ASF, SRp40, SRp55 and SC35. Sequences matching these consensus ESEs stimulate splicing of conventional and non-conventional introns (Liu et al., 1998).

Dr. Liu also studied a nonsense mutation in BRCA1, the breast cancer susceptibility gene, that is associated with familial breast and ovarian cancer. This mutation causes skipping of the entire exon (Mazoyer et al., 1998). Dr. Liu found that this mutation disrupts an SF2/ASF high score recognition motif which may function as an exonic splicing enhancer. *In vitro* splicing analysis showed that exon skipping is due to the failure to recognize this ESE.

I took over the project in May, 1999 when Dr. Liu left the laboratory to accept a research position at Phylos Inc. I investigated the role of SR proteins in splicing of AT-AC introns. Little is known about the role of conventional splicing factors in the excision of AT-AC introns. The two types of introns are spliced by homologous but distinct spliceosomes (Tarn and Steitz, 1996; Wu and Krainer, 1996). Similar to conventional introns, AT-AC intron splicing is stimulated by exon definition interactions mediated by exonic splicing enhancers. The study of AT-AC intron splicing allows comparison of splicing in the two types of splicing and will thereby provide insight into the requirements for activity of specific splicing factors. I found that individual SR proteins, which are required for conventional intron splicing, are also required for excision of a non-conventional, AT-AC intron in the SCN4A gene. SR protein-dependent enhancement of AT-AC intron splicing was also observed in the presence of an exonic splicing enhancer or downstream 5'ss, similar to the activity in major introns.

(6) Body

SR proteins recognize specific exonic splicing enhancer sequences

Exonic splicing enhancer sequences recognized by human SR proteins, SF2/ASF, SRp40, SRp55, and SC35 were identified by performing a randomization and selection procedure under splicing conditions (Liu et al., 1998, 2000). The selected sequences are functional in splicing enhancer activity and are specific for the individual SR protein which was used in the selection.

BRCA1 mutation disrupts exonic splicing enhancer

A large number of gene function defects and human diseases are attributed to point mutations. Point mutations in a gene may alter activity of an important protein domain or create nonsense mutations that would activate an RNA decay pathway. Silent or missense mutations as well as nonsense mutations have been reported to cause exon skipping. One explanation for these results would be the disruption of exonic splicing enhancer sequences which are recognized by specific SR proteins during splicing. The discovery of exonic sequences which are critical for correct pre-mRNA splicing leads to the hypothesis that mutations which previously were assumed to affect protein activity or RNA stability may actually cause splicing defects. Further analysis of this hypothesis will have important implications for therapeutic approaches for treatment of human disease and cancer. Human gene mutations were searched using the score matrices for SF2/ASF, SC35, SRp40 and SRp55 which were generated from the consensus sequences of selected enhancers. Dr. Liu searched 16 nonsense mutations, 13 missense mutations and two silent mutations. These mutations had been previously reported to cause exon skipping. Remarkably, the search showed that most of the mutations eliminate one or more high score motifs of specific SR proteins. The conclusion is that many point mutations could cause exon skipping by disrupting exonic splicing enhancers. To directly test the ability of a human gene point mutation to disrupt splicing by altering an ESE, Dr. Liu analyzed splicing of a mutant form of the breast cancer susceptibility gene, BRCA1, which results in exon skipping (Mazoyer et al, 1998). A nonsense mutation (G to T) of Glu1694 in exon 18 of BRCA1 was found in a family with four cases of breast cancer and four cases of ovarian cancer. The skipping of exon 18 retains the same reading frame and removes 26 amino acids of BRCA1. Score matrices of SF2/ASF, SC35, SRp40 and SRp55 ESEs were used to search the wild type and mutant BRCA1 gene. We found that an SF2/ASF high score motif in exon 18 was disrupted by the BRCA1 point mutation. To investigate the

mechanism for exon skipping activated by this mutation, wild type and mutant BRCA1 minigenes spanning from exon 17 to exon 19 were constructed and used in in vitro splicing assays in HeLa nuclear extract. Dr. Liu found that in the wild type minigene exon 18 was always spliced and therefore included in the mRNA whereas the a large percent of the mutant exon 18 was skipped and not spliced. This result supports the findings from the search with SR protein high score motifs and suggests that the mutation disrupts an exonic splicing enhancer. The nonsense mutation in BRCA1 may function to activate nonsense-mediated decay of the RNA and the exon skipping itself may be a component of this yet unknown decay path but not the causative defect in BRCA1 gene expression. To test whether disruption of the splicing enhancer and not introduction of a codon mutation causes the BRCA1 altered splicing, Dr. Liu constructed two additional minigenes. One, LM, has a missense mutation but no SF2/ASF high score motif. The second mutant, HM, has a nonsense mutation but maintains an SF2/ASF high score motif. In vitro splicing assays of these mutants showed a high ratio of exon 18 inclusion relative to skipping in the splicing of the HM nonsense codon mutant. In contrast skipping of exon 18 was relatively high in the missense mutant LM. This result indicates a strong correlation between exon skipping and disruption of the SF2/ASF high score motifs.

SR proteins are required for AT-AC intron splicing

AT-AC intron excision occurs by a two-step splicing pathway similar to that of conventional GT-AG introns. Despite this similarity, most of the snRNA components of the AT-AC spliceosome are distinct from those of the conventional GT-AG spliceosome. This difference likely reflects a need for altered specificity to preserve key RNA:RNA interactions at the splice sites. Whether protein components of the AT-AC spliceosome are novel or shared between the two spliceosomes is not known. SR proteins play important roles in several aspects of splicing of GT-AG introns in higher eukaryotes. To assess a potential role for SR proteins in AT-AC splicing, an assay was developed using HeLa cell S100 extract. Conventional introns are spliced in S100 extract only if SR proteins are also added. Splicing of the SCN4A pre-mRNA AT-AC intron 2 as well as the E2F4 AT-AC intron 3 was not detected in this assay unless an additional nuclear extract fraction was included. In the presence of this fraction and the S100 extract, a dependence on total HeLa SR proteins was observed. AT-AC splicing also was observed using individual recombinant SR proteins, SF2/ASF and SRp55 while SC35 was not able to activate AT-AC intron splicing. This result demonstrates that conventional SR proteins are required for excision of AT-AC introns. Multiple individual SR proteins could

promote the splicing reaction, suggesting that SR proteins are functionally redundant in AT-AC splicing. AT-AC intron splicing is stimulated over basal splicing by exon-definition interactions with a downstream conventional 5′ splice site or by an exonic splicing enhancer. S100 complementation assays indicate that SR proteins stimulate enhancer- and downstream 5′ splice site-dependent splicing as well as basal AT-AC splicing. Whether different SR proteins are required for basal splicing and enhancer or downstream 5′ splice site-dependent splicing is being investigated.

References

Ghigna, C., Moroni, M., Porta, C., Riva, S., and Biamonti, G. (1998) Altered expression of heterogeneous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. Canc. Res. 58:5818-5824.

Liu, H-X., Chew, S.L., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (1999) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol. Cell. Biol. *in press*.

Liu, H-X, Zhang, M. and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes & Dev. 12:1998-2012.

Mazoyer, S., Puget, N., Perrin-Vidoz, L., Lynch, H.T., Serova-Sinilnikova, O.M., and Lenoir, G.M. (1998) A BRCA1 nonsense mutation causes exon skipping. Am. J. Hum. Genet. 62:713-715.

Stickeler, E., Kittrell, F., Medina, D., and Berget, S.M. (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. Oncogene 18:3574-3582.

Tarn, W-Y. and Steitz, J.A. (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vivo. Cell 84:801-811.

Wu, Q. and Krainer, A.R. (1996) U1-mediated exon definition interactions between AT-AC and GT-AG introns. Science 274:1005-1008.

(7) APPENDIX

1) KEY RESEARCH ACCOMPLISHMENTS

- Consensus sequences of exonic splicing enhancers (ESEs) recognized specifically by individual SR proteins, SF2/ASF, SRp55, SRp40, and SC35, were identified by an in vitro SELEX procedure under splicing conditions.
- A nonsense mutation in BRCA1, the breast cancer susceptibility gene, that is associated with familial breast and ovarian cancer causes skipping of the entire exon. This mutation is within a consensus SF2/ASF recognition site and *in vitro* splicing analyses showed that exon skipping is due to disruption of the putative SF2/ASF exonic splicing enhancer sequence.
- SR proteins, which are required for conventional intron splicing, are also required for excision of non-conventional, AT-AC introns.

2) REPORTABLE OUTCOMES

publications

Liu, H-X., Chew, S.L., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (1999) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol. Cell. Biol. *in press*.

Liu, H-X., Zhang, M. and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes & Dev. 12:1998-2012.

presentations

Hastings, M.L. and Krainer, A.R. (1999) AT-AC intron splicing requires SR proteins. poster presentation. Eukaryotic mRNA processing meeting. Cold Spring Harbor Laboratory, August, 1999.

Hastings, M.L. and Krainer, A.R. (1999) SR proteins are required for AT-AC intron splicing. oral presentation. RNA99 meeting. Edinburgh, Scotland, June, 1999.

employment

Dr. Liu is presently employed at Phylos Inc., 128 Spring St., Lexington, MA 02421

Exonic splicing enhancer motif recognized by human SC35 under splicing conditions

Hong-Xiang Liu,[†] Shern L. Chew,[‡] Luca Cartegni, Michael Q. Zhang, and Adrian R. Krainer[*]

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724-2208

Running title: SC35-Specific Exonic Splicing Enhancers

[*]Corresponding author

[†]Present address: Phylos Inc., 128 Spring Street, Lexington, MA 02421

[‡]Present address: Department of Endocrinology, St Bartholomew's and the Royal London School of Medicine and Dentistry, London EC1A 7BE, UK

Editorial correspondence to:

Dr. Adrian R. Krainer
Cold Spring Harbor Laboratory
1 Bungtown Road, PO Box 100
Cold Spring Harbor, NY 11724-2208
Tel: (516) 367-8417
Fax: (516) 367-8453
E-mail: krainer@cshl.org

ABSTRACT

Exonic splicing enhancers (ESEs) are important cis-elements required for exon inclusion. Using an in vitro functional selection and amplification procedure, we have identified a novel ESE motif recognized by the human SR protein SC35 under splicing conditions. The selected sequences are functional and specific: they promote splicing in nuclear extract, or in S100 extract complemented by SC35 but not by SF2/ASF. They can also function in a different exonic context from the one used for the selection procedure. The selected sequences share one or two close matches to a short and highly degenerate octamer consensus, GRYYcSYR. A score matrix was generated from the selected sequences according to the nucleotide frequency at each position of their best match to the consensus motif. The SC35 score matrix, along with our previously reported SF2/ASF score matrix, was used to search the sequences of two well characterized splicing substrates derived from the mouse IgM and HIV tat genes. Multiple SC35 high-score motifs, but only two widely separated SF2/ASF motifs, were found in the IgM C4 exon, which can be spliced in S100 extract complemented by SC35. In contrast, multiple high-score motifs for both SF2/ASF and SC35 were found in a variant of the Tat T3 exon (lacking an SC35-specific silencer) whose splicing can be complemented by either SF2/ASF or SC35. The motif score matrix can help locate SC35-specific enhancers in natural exon sequences.

INTRODUCTION

Accurate removal of introns from pre-mRNA requires multiple cis-elements, including the splice sites, polypyrimidine tract, branch site, and other intronic and exonic sequences that have positive or negative effects on splicing (10, 32, 45; reviewed in references 1 and 2). Positive-acting sequences, termed exonic splicing enhancers (38, 40, 43), have been identified primarily in exons associated with regulated splicing. These exons are typically adjacent to introns with weak intronic splicing signals and require ESEs for their inclusion. Deletion of an ESE often causes exon skipping, or in the case of terminal exons, suppresses removal of the last intron. One of the first characterized ESEs is located in the M2 3'-terminal exon of the mouse IgM gene (40). This 73-nt ESE, which is highly purine rich, is required for inclusion of the alternatively spliced M2 exon. However, deletion of just the purine-rich sequences within this ESE does not abolish splicing completely. The M2 ESE also functions in a heterologous context to enhance splicing of a *Drosophila* doublesex intron (40).

A SELEX procedure has been used to identify sequences that can function as ESEs (37). A 20-nt sequence of the internal duplicated exon of a model pre-mRNA was replaced by 20 nt of random sequence. The randomized pre-mRNAs were incubated under splicing conditions in nuclear extract and functional enhancer elements that promoted splicing were selected. A large number of sequences, both purine rich and non-purine rich were obtained, and the two types of sequences stimulated exon inclusion to similar extents. A similar approach was used in an in vivo system involving transfection of a troponin minigene with random sequences in place of a natural ESE (9). Purine-rich sequences and a novel class of AC-rich ESE sequences were identified. The AC-rich sequences are efficient splicing enhancers and can also function in a heterologous gene context.

Considerable evidence suggests that ESEs interact specifically with a family of RNA-binding proteins called SR proteins, which are characterized by one or two RNA-recognition motifs (RRMs) and a C-terminal Arg/Ser-rich domain (12, 18, 28, 35, 38, 39). SR proteins are essential splicing factors required for both constitutive and alternative splicing (11, 17, 44). SR proteins can determine alternative splice-site selection by antagonizing the activity of hnRNP A/B proteins. High concentrations of SR proteins usually favor the use of proximal splice sites and exon inclusion, whereas high concentrations of hnRNP A/B proteins tend to favor distal splice sites and exon skipping (5, 23, 26). SR proteins also specifically recognize ESEs, and the resulting complex may then recruit U2AF binding to the weak polypyrimidine tract of the upstream 3' splice site. The ESE-SR protein-U2AF interaction is thought to be important during the early stages of spliceosome assembly (8, 16, 42, 46), although recent evidence suggests that in at least some cases, including the IgM M2 exon, ESEs act in part by neutralizing exonic silencer elements (7, 15). SF2/ASF and SC35 are two of the best characterized among the nine human SR proteins identified to date. Both proteins have been implicated in many aspects of constitutive and regulated splicing. Both are found in the pre-spliceosomal E complex and can interact with U1-70K and U2AF by RS-domain-mediated protein-protein interactions. The RRMs of these two proteins are responsible for their unique substrate specificities (6, 27).

A better understanding of the functional interactions between ESEs and SR proteins depends on knowledge of the sequence specificity of all SR proteins. To this end, we recently performed an iterative selection under splicing conditions to identify exon sequences that can enhance splicing in the presence of each of three SR proteins. We identified three novel classes of functional ESE motifs recognized specifically by SF2/ASF, SRp40, and SRp55. The consensus motifs indicated that individual SR proteins recognize distinct and highly degenerate sequences (20). The three SR proteins we studied previously are closely related, i.e., they all have two tandem RRMs. To extend

this analysis, we have now determined the sequence specificity of an additional, extensively studied SR protein, SC35, which has a single, N-terminal RRM.

MATERIALS AND METHODS

Preparation of Hela cell extract and recombinant SR proteins. Hela nuclear and S100 extracts were prepared as described (24). Recombinant SC35 expressed in baculovirus was generously provided by K. Lynch and T. Maniatis, and by R.-M. Xu.

Selection/amplification procedure. The amplification and selection procedure was carried out as described (20). Briefly, the natural ESE of the IgM pre-mRNA was replaced by 20 nt of random sequence using overlap-extension PCR with μMΔ DNA (40) as a template. The resulting PCR product was used for in vitro transcription to generate a ^{32}P-labeled random pre-mRNA pool. 20 fmol of the pre-mRNA pool was incubated under in vitro splicing conditions in S100 extract plus recombinant SC35 in a 25 μl reaction. The RNA was separated by denaturing PAGE, and the spliced mRNAs were excised and eluted from the gel in 0.5 M ammonium acetate + 0.1% SDS, and re-amplified by RT-PCR. Reverse transcription was carried out using Superscript II as described by the manufacturer (Life Technologies). PCR was performed using high-fidelity Pfu polymerase as specified by the manufacturer (Stratagene). The PCR product was subcloned into the vector PCR-Blunt (Stratagene) and sequenced using a Dye Terminator Cycle Sequencing kit (Perkin-Elmer) and an automated ABI 377 sequencer. Selected winner sequences were rebuilt into DNA templates for transcription of pre-mRNAs by overlap-extension PCR, as done initially for the random sequences (20).

Sequence analysis and construction of score matrices. The selected sequences of each SR protein winner pool plus a portion of the flanking nucleotides were aligned using Gibbs sampler (19). The identified consensus motif was then used to generate a

score matrix. The compositional bias of the initial RNA pool was taken into account. For details of the sequence analysis, see (20).

In vitro splicing. PCR products carrying an SP6 or T7 promoter were used for in vitro transcription. 5' capped transcripts were incubated in 25-μl splicing reactions as described (25). Each reaction had 4 μl of nuclear extract or 7 μl of S100 extract. For S100 complementation assays, 20 pmol of specific SR protein was used. Splicing reactions were carried out at 30 °C for 4 hr. The RNA was then extracted, loaded on 6% or 12% polyacrylamide gels, and visualized by autoradiography (20). DNA templates for IgM M1-M2 pre-mRNAs with a D2 variant containing an SC35 consensus match or with the 6-24 winner sequence (30) were made by overlap-extension PCR using primers M2-D2HXL (GTGAAATGACTCTCAGCATggggacatactcggcccctgCTAGTAAACTTATTC-TTACGT) and M2-SCH24 (GTGAAATGACTCTCAGCATtttgcggtctccggcctccCTAGT-AAACTTATTCTTACGT), respectively. DNA templates for pre-mRNAs in an IgM C3-C4 context were made by PCR on pμC3-C4 plasmid DNA (41) using an SP6 promoter primer and the following antisense primers: Ca (TGGCAGCAGGTACACAGC); CaCb (gtggctgactccctcagg); D2 (ctgcggccgagtatgtccccTGGCAGCAGGTACACAGC); D2C (c-aggggccgagtatgtccccTGGCAGCAGGTACACAGC); and 6-24 (ggaggccggagaccgcaaaT-GGCAGCAGGTACACAGC). RNAs were made as described above.

RESULTS

Identification of ESE motifs recognized by SC35 under splicing conditions. To study the sequence specificity of ESE recognition by SC35 under splicing conditions, a functional SELEX procedure (20) was used (Fig. 1). Functional ESEs were selected in the context of a well-characterized mouse immunoglobin μ heavy chain minigene transcript, comprising the last intron flanked by the M1 and M2 exons (40). The natural ESE in the M2 exon was replaced by 20 nt of random sequence using overlap-extension

PCR. The random RNA pool, a library of pre-mRNAs representing 1.2 x 10^{10} different molecules, was spliced in nuclear extract or in S100 extract complemented by SC35. As previously reported, the wild-type IgM pre-mRNA spliced very efficiently in nuclear extract, with the mature mRNA representing greater than 90% of the RNA after a 4-hr incubation (Fig. 2, lane 1). In contrast, the ESE-deleted mutant ED did not splice at all under the same conditions (Fig. 2, lane 2) confirming that the natural ESE of the IgM pre-mRNA is essential for splicing (40). The initial RNA pool was spliced in nuclear extract with an apparent efficiency of about 20% (Fig. 2, lane 3), whereas no splicing was detected in the S100 extract alone (lane 4). When the S100 extract was complemented by SC35, splicing of the initial RNA pool remained undetectable by autoradiography (lane 5). However, we assumed that a very small fraction of the RNA pool was correctly spliced, and we excised a gel slice corresponding to the position of spliced mRNA, using the product in lane 3 as a marker. RNA was eluted from the gel slice and amplified by RT-PCR. The amplified products were cloned and 30 clones were sequenced. The resulting sequences were analyzed using the program GIBBS sampler to determine a consensus sequence (19, 20). A score matrix was generated according to the frequency of each nucleotide at each position of the consensus motif, adjusted for the compositional bias of the initial random pool. This score matrix was used to identify high-score motifs within each winner sequence, taking into account the randomized sequence region and a small portion of the flanking sequences.

The SC35 winner sequences after a single round of selection yielded the short degenerate octamer consensus motif GRYYcSYR (Fig. 3). The C-residue content within the randomized 20-nt segment increased from 19% in the initial pool to 23% after a single round of selection. This change in C composition occurred at the expense of slight reductions in the content of G, A, and U residues. As reported previously for our similar analysis of other SR proteins, the SC35 consensus sequence is highly degenerate. Several of the winner sequences have more than one high-score motif (Fig. 3A). The scores of the

30 SC35 winner sequences range from 1.19 to 3.55, with a mean score of 2.56 ± 0.56. 30 individual sequences cloned and randomly selected from the initial pool (20) gave a range of scores from 0.64 to 3.23, with a mean score of 1.62 ± 0.72 when searched by the same score matrix. Only 3 sequences in the control pool had scores higher than the mean of the winner pool, whereas 16 sequences in the winner pool had scores higher than this mean, and 28 had scores higher than the mean of the control pool. The difference in the means of the scores between the two sequence pools is highly significant (p-value $< 10^{-7}$ under a t-test with df = 58).

The highest possible score for a single octamer is 3.95, corresponding to the sequence GGCCCCUG (Fig. 3B). This precise sequence does not occur in any of the 30 winner sequences analyzed. The absence of a "perfect" motif in the selected sequences may reflect the small sample size, or the fact that a linear consensus sequence or nucleotide frequency matrix assumes an independent contribution at each position, an assumption that may or may not fit the actual recognition mechanism (34).

The selected SC35 sequences are functional and specific ESEs. We next tested whether the individual sequences selected in the presence of SC35 could function as true splicing enhancers. Five sequences with a range of scores were arbitrarily chosen from the 30 analyzed sequences and individually rebuilt into IgM M1-M2 pre-mRNAs with the same structure as those in Figs. 1 and 2, using overlap-extension PCR and in vitro transcription (20). Each pre-mRNA was then incubated under splicing conditions in nuclear extract or in S100 extract complemented by SC35 (Fig. 4). Four of the five SC35 winner sequences activated IgM pre-mRNA splicing very efficiently in nuclear extract (Fig. 4, lanes 1, 4, 7, and 10). They also promoted IgM pre-mRNA splicing in S100 extract complemented by SC35, albeit less efficiently (lanes 3, 6, 9, and 12), but not in S100 extract alone (lanes 2, 5, 8, and 11). One winner sequence from the SC35 winner pool, D5, enhanced splicing less efficiently in nuclear extract (lane 13), and gave only trace activity in the complementation assay (lane 15). In general, the splicing efficiency

correlated with the motif scores shown in Fig. 3. D1 and D2 have the highest scores; D3 and D4 have intermediate scores; and D5 has the lowest score among the 30 sequences analyzed (Fig. 3). However, the correlation between splicing efficiency and motif scores is not linear, presumably reflecting sequence context effects. Also, D3 has a higher score than D4, and although they spliced with similar efficiency in nuclear extract, D4 spliced more efficiently in the complementation assay. 16 sequences from the random RNA pool were also analyzed for enhancer activity (20). All of them spliced in nuclear extract poorly or not at all. In most cases the pre-mRNAs showed partial degradation, suggesting that spliceosomal complexes did not assemble on these RNAs (21).

Next, we determined the SR protein specificity of the SC35-selected ESEs. Pre-mRNAs with the different winner sequences were separately incubated under splicing conditions in S100 extract complemented by SC35, SF2/ASF, SRp40 or SRp55. All of the tested SC35 winners promoted splicing with higher efficiency in S100 extract when the extract was complemented by SC35 (Fig. 5A, lanes 3, 7, 11, 15 and 19), SRp40 or SRp55 (21). When the extract was complemented by SF2/ASF, the splicing efficiencies were much lower (Fig. 5A, lanes 4, 8, 12, 16 and 20). In contrast, five SF2/ASF-selected winners promoted splicing in S100 extract complemented by either SF2/ASF (Fig. 5B, lanes 4, 8, 12, 16 and 20; reference 20) or SC35 (Fig. 5B, lanes 3, 7, 11, 15, and 19) with comparable efficiencies. These SF2/ASF winners promoted splicing very poorly or not at all in the presence of SRp40 or SRp55 (20).

Comparison of SC35 ESE motifs and activity in different exonic contexts. To test whether an octamer with the highest possible SC35 ESE score has enhancer activity and to compare this consensus with a previously identified one, we analyzed the D2 winner containing the motif GGCCGCAG, a variant of D2 with two transversions that create the maximum score consensus GGCCCCUG, and one of the 19mer winners (6-24) selected by Schaal and Maniatis (30). These three sequences were first tested in the context of the IgM M2 exon (Fig. 6A). All three sequences strongly promoted splicing of

exons M1 and M2 in nuclear extract (lanes 3-5), in contrast to the lack of detectable splicing with the parent pre-mRNA in which the natural ESE was deleted (lane 1).

Next we tested the same three ESEs in a different exonic context, namely the C4 exon derived from a different region of the IgM pre-mRNA. When this exon is divided into three segments, Ca, Cb, and Cc, the Cc segment is dispensable, whereas the Cb segment behaves as an SC35-specific ESE (27). Indeed, a shortened 3' exon consisting of the Ca and Cb segments of C4 spliced to exon C3 much more efficiently in nuclear extract than one consisting of Ca alone (Fig. 6B, lanes 1 and 2). When the Cb segment was replaced by each of the above three ESEs, all of them promoted splicing above the background of Ca alone (lanes 4-6). However, the D2 winner ESE was as strong as the natural Cb ESE, the 6-24 ESE was slightly less efficient, and the perfect consensus was the least active. These results show that both our SC35 motif and a winner sequence identified in a previous study (30) can function in different exonic contexts, although the precise context can influence the extent of enhancement.

Distribution of SC35 ESE motifs in natural genes. To determine whether the selected ESE motifs are relevant to splicing of natural pre-mRNA substrates, we conducted a search of SC35 high-score motifs in natural genes. Only scores higher than the lowest score of the SC35 winner pool are shown (Fig. 7, green vertical bars). For comparison, we also show the high-score SF2/ASF motifs in the same genes (Fig. 7, blue vertical bars; reference 20). The first natural sequence we examined was the M2 exon of the IgM gene. The search result indicated that there are many SC35 ESE motifs within the segment comprising the previously characterized natural ESE (Fig. 7A, magenta horizontal bar). The distribution of high-score SC35 motifs differs from that of SF2/ASF motifs. SF2/ASF-specific motifs are present at a higher density within the natural ESE than in the flanking regions. In contrast, high-score SC35 motifs have a relatively even distribution across the M2 exon. Both SR proteins can promote splicing of this pre-mRNA in S100 extract (21). The presence of ESE motifs in regions lacking enhancer

activity shows that although the motifs may be necessary, they are not sufficient for ESE function (see Discussion).

To address the issue of whether the identified ESE motifs are specific to SC35, we searched two additional pre-mRNA substrates that are known to have different SR protein specificities. Splicing of the IgM C3-C4 pre-mRNA is activated in S100 extract when complemented by SC35 but not by SF2/ASF (27). In contrast, splicing of the HIV Tat T2-T3 pre-mRNA is activated by SF2/ASF but not by SC35 in S100 extract (6, 27). When an SC35-specific splicing silencer in the 3' region of the T3 exon is deleted, both SF2/ASF and SC35 can activate T2-T3 splicing in S100 extract. Detailed analysis of the splicing of these two pre-mRNAs indicated that the C4 and T3 exons determine the SR protein specificity (27). Our search result matches the experimental data (Fig. 7B and C). Many high-score motifs matching the consensus of SC35 were found in the C4 exon, but only two well-separated SF2/ASF motifs were found in this exon (Fig. 7B). Interestingly, in a deletion mutant missing the first 38 nt of the C4 exon, splicing of C3-C4 was activated by both SF2/ASF and SC35 (27). Consistent with this result, the SF2/ASF motif near position 61 is closer to the 3' splice site in the deletion mutant. High-score motifs for both SF2/ASF and SC35 were found in the T3 exon of the Tat gene (Fig. 7C). Curiously, a single SC35 high-score motif is present within the SC35-specific silencer region.

Finally, we studied the distribution of SC35 high-score motifs in human exons versus introns. A total of 570 genes, representing 2,626 exons (426 kb) and 2,079 introns (1,295 kb), were extracted from the ALLSEQ database (4) and analyzed. Scores equal to or higher than the mean score of the winner pool were taken into account. High-score motifs appeared more frequently in exons than in introns. An average of 9 SC35 high-score motifs were found per kilobase of exon compared to only 5.9 per kilobase of intron. This comparison was statistically significant because of the large database size (p-value < 10^{-10}).

DISCUSSION

We have identified a novel ESE motif recognized by the human SR protein SC35. Several lines of evidence point to the biological relevance of the selected ESE motifs. First, they are functional ESEs. All of the SELEX winners we have tested promote splicing in nuclear extract, and in S100 extract plus the cognate SR protein. In nuclear extract, the SELEX winners function as potent ESEs. Second, the SC35 motifs are present within exon segments containing natural ESEs, and are more frequently found in exons than in introns, suggesting that they may contribute to exon definition by the spliceosome. Third, the SC35 motifs are specific, i.e., they are not recognized by all SR proteins. The SC35-selected ESEs were recognized by SC35, SRp40 or SRp55, but not by SF2/ASF under splicing conditions. In addition, the distribution of high-score motifs of SF2/ASF and SC35 in the IgM C4 and Tat T3 exons correlated with the observed SR protein specificity of the corresponding substrates (27). This result also suggests that the score matrices we have generated have some predictive value. We have previously analyzed the predictive value of the SR-specific score matrices derived for other SR proteins (20). Statistically, high-score motifs of SR proteins are present at a higher density in natural ESEs than in the flanking regions. Experimentally, SR proteins specifically recognize their cognate ESE motifs when these are placed in the context of the IgM M2 exon, replacing the natural ESE. The present study confirms and extends our previous work to two natural ESEs in IgM and HIV tat exons. In addition, we have now shown that SC35 winner sequences and a maximum-score SC35 motif can promote splicing in different exonic contexts.

The specific interaction between SR proteins and ESEs has also been described in other systems. During assembly of enhancer complexes in vitro (Enh complex, resembles the E complex), the enhancer sequences determine the specific pattern of SR proteins that can be UV crosslinked to the RNA (33). Female-specific alternative splicing of the

Drosophila doublesex pre-mRNA requires six 13-nucleotide repeat elements (dsxRE) and a purine-rich element (PRE). UV cross-linking analysis showed that SR proteins, along with Tra and Tra-2, assemble on the ESEs in a stepwise and sequence-specific manner (22). The fact that SR proteins are expressed in a tissue-specific manner (14, 44), together with the specific recognition of ESEs by individual SR proteins may contribute quantitatively to the regulation of gene expression.

The SC35 SELEX winners have the consensus GRYYcSYR, which is a highly degenerate sequence. Even though SC35 has a single RRM, a SELEX protocol based on RNA binding yielded two different nonamer consensus sequences, AGSAGAGTA and GTTCGAGTA, which share the last five nucleotides (36). These two motifs differ significantly from the more degenerate consensus identified by functional SELEX. Although the second motif has a partial fit to the above consensus, neither motif has a good score, consistent with the observation that the high-affinity binding sequences fail to enhance splicing of RNA substrates in nuclear extract or in S100 extract plus SC35, even when present in several copies (36). Therefore, it appears that high-affinity SC35-binding sites are not optimal for function. Perhaps RNA-binding selection does not achieve an interaction geometry compatible with SC35 enhancement function, or it is essential to co-select sequences that in addition to binding SC35 can also accommodate putative co-activators or fail to bind silencing factors.

Nevertheless, our data argue that SC35 has limited, but defined sequence specificity in recognizing functional sequences. Despite the fact that this protein has a single RRM, the functional recognition motif is degenerate, as was the case for the two-RRM SR proteins SF2/ASF, SRp40 and SRp55 (20). Therefore, the degeneracy of the ESE motifs recognized by those proteins is probably not attributable to the recognition of distinct motifs by each of their RRMs. The sequence degeneracy of the ESEs is consistent with the fact that they must coexist with a very wide variety of unrelated open reading frames and must be recognized by a discrete set of SR proteins (20, 29, 30).

Schaal and Maniatis recently used a similar functional SELEX approach to select ESEs that could function in the context of the *Drosophila* doublesex pre-mRNA in Hela nuclear extract (30). The selected 18-nt winner sequences were then individually analyzed by S100 complementation assays to define their SR protein specificity. Two round-6 winner sequences were the most active in the presence of SC35. By comparing these two sequences to each other and to an SC35-dependent ESE present in human β-globin exon 2, the authors proposed the SC35 heptamer consensus UGCNGYY, which is also a highly degenerate sequence. Although this heptamer motif is substantially different from our consensus octamer motif, some versions of the degenerate heptamer consensus have high scores, as defined in the present study. We therefore searched the two published winner sequences (30) using our SC35 score matrix. Both sequences had multiple high-score motifs, some of which were non-overlapping, consistent with the fact that they had undergone six rounds of selection for splicing. In the case of the 6-24 sequence, the highest score (3.13) corresponded to the octamer GGUCUCCG, which has a 4-nt overlap with the UGCGGUC sequence that fits the heptamer consensus. In the case of the 6-38 sequence, the second highest score (1.56) corresponds to the octamer UGCCGCCG, of which the first seven nucleotides fit the heptamer consensus; the highest score (2.44) was for the non-overlapping octamer GGACCGGA. Similarly, within the 18-nt β-globin fragment in which Schaal and Maniatis characterized an SC35-dependent ESE that comprises the heptamer UGCUGUU (29), the highest score (1.36) corresponds to the octamer UGAUGCUG, which includes the first five nucleotides of the heptamer.

We conclude that despite the very different pre-mRNA contexts, types of extract used for the selection, and number of selection rounds, the SC35 ESEs identified by the two approaches are remarkably consistent. We believe, however, that our octamer motif has greater predictive value because it was derived from a much larger number of winner sequences. Also, the use of a nucleotide frequency matrix derived from 30 sequences allows identification of putative SC35 ESEs that do not precisely match the consensus at

every position. Thus, our SC35 score matrix finds high-score motifs in both of the winner sequences and the β-globin segment characterized by Schaal and Maniatis (29, 30), whereas of our 30 SC35 winner sequences (Fig. 3), only number 14 has a precise match to the heptamer consensus they defined.

The IgM M2 exon has a higher density of SF2/ASF and SRp40 high-score motifs within the natural ESE segment than in the flanking sequences. In contrast, the SRp55 high-score motifs do not correlate with the location of the ESE (20). In the case of SC35, the high-score motifs also have a relatively even distribution across the exon. The different motif distributions may reflect different mechanisms of SR protein-ESE recognition. Although for some pre-mRNAs any SR protein can complement splicing in the S100 extract (31, 44), each SR protein may function by slightly different mechanisms. Some SR proteins may require multiple binding sites to function, and the optimal distance from the 3' splice site to the SR protein-binding site may also be protein specific. The fact that ESE motifs are not found exclusively in natural exonic segments required for ESE activity indicates that the motifs are not sufficient for ESE function. It appears that sequence context, structure, or position effects are also very important.

Examples of sequence context effects that can influence ESE activity are provided by exonic splicing silencers. These inhibitory elements probably co-exist with splicing enhancers in many exons, and they may also be SR protein-dependent and function in a cell-type specific manner. For example, an SC35-dependent silencer sequence has been mapped in the Tat gene T3 exon (27). This silencer element includes within it an SC35-specific ESE motif (Fig. 7C). We speculate that binding of SC35 to this region prevents the function of other splicing factors, although it is presently unclear how this element acts at a distance and suppresses the effect of SC35-dependent ESEs but not of SF2/ASF-dependent ESEs. Recently, the 3' portion of the IgM M2 exon was also shown to comprise a silencer element that binds U2 snRNP and antagonizes the upstream ESE (15). The silencer element, so far mapped to a fragment between nucleotides 94-167 (see

Fig. 7A) in the M2 exon, overlaps with several SC35 high-score motifs and with one SF2/ASF high-score motif.

The similar arrangement of adjacent ESE and ESS elements seen in the IgM M2 and Tat T3 exons may turn out to be a common feature of many vertebrate cellular and viral exons. To improve the predictive value of the SR-protein-specific ESE motifs, it will be necessary to gain a better understanding of the influence of sequence context and position, as well as of the mechanistic basis for the function of splicing enhancers and silencers.

ACKNOWLEDGMENTS

We thank Prof. Y. Shimura for the gift of IgM plasmids, Drs. K. Lynch, T. Maniatis, R.-M. Xu and A. Mayeda for recombinant SR proteins, J. Yin for DNA sequencing, A. Mayeda and members of our laboratory for valuable ideas, and M. Hastings for helpful comments on the manuscript.

This work was supported by NIH grants to A.R.K. (GM42699) and M.Q.Z. (HG01696), by an Advanced Fellowship from The Wellcome Trust to S.L.C. (045401), by a fellowship from the Human Frontiers Science Program to L.C. (LT0066/1997-M), and by a fellowship from the U.S. Army Medical Research and Matériel Command under DAMD 17-96-1-6172 to H.-X.L.

REFERENCES

1. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. **270:**2411-2414.
2. **Black, D. L.** 1995. Finding splice sites within a wilderness of RNA. RNA **1:**763-771.

3. **Burge, C. B., T. Tuschl, and P. A. Sharp.** 1999. Splicing of Precursors to mRNAs by the Spliceosomes, p. 525-559. *In* R. F. Gesteland, T. R. Cech, and J. F. Atkins (ed.), The RNA World, Second Edition. Cold Spring Harbor Laboratory Press, New York.

4. **Burset, M., and R. Guigo.** 1996. Evaluation of gene structure prediction programs. Genomics **34:**353-367.

5. **Cáceres, J. F., S. Stamm, D. M. Helfman, and A. R. Krainer.** 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science **265:**1706-1709.

6. **Chandler, S. D., A. Mayeda, J. M. Yeakley, A. R. Krainer, and X.-D. Fu.** 1997. RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc. Natl. Acad. Sci. USA **94:**3596-3601.

7 **Chew, S. L., H.- L. Liu, A. Mayeda, and A. R. Krainer.** 1999. Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. Proc. Natl. Acad. Sci. USA **96:**10655-10660.

8. **Chiara, M. D., and R. Reed.** 1995. A two-step mechanism for 5' and 3' splice-site pairing. Nature **375:**510-513.

9. **Coulter, L. R., M. A. Landree, and T. A. Cooper.** 1997. Identification of a new class of exonic splicing enhancers by in vivo selection. Mol. Cell. Biol. **17:**2143-2150. (Erratum, **17:**3468.)

10. **Freyer, G. A., J. P. O'Brien, and J. Hurwitz.** 1989. Alterations in the polypyrimidine sequence affect the in vitro splicing reactions catalyzed by HeLa cell-free preparations. J. Biol. Chem. **264:**14631-14637.

11. **Ge, H., P. Zuo, and J. L. Manley.** 1991. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. Cell **66:**373-382.

12. **Gontarek, R. R., and D. Derse.** 1996. Interactions among SR proteins, an exonic splicing enhancer, and a lentivirus Rev protein regulate alternative splicing. Mol. Cell. Biol. **16:**2325-2331.

13. **Gorodkin, J., L. J. Heyer, S. Brunak, and G. D. Stormo.** 1997. Displaying the information contents of structural RNA alignments: the structure logos. Comput. Appl. Biosci. **13:**583-586.

14. **Hanamura, A., J. F. Cáceres, A. Mayeda, B. R. Franza, Jr., and A. R. Krainer.** 1998. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA **4:**430-444.

15. **Kan, J. L., and M. R. Green.** 1999. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. Genes Dev. **13:**462-471.

16. **Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Lührmann, M. A. Garcia-Blanco, and J. L. Manley.** 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature **368:**119-124.

17. **Krainer, A. R., A. Mayeda, D. Kozak, and G. Binns.** 1991. Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. Cell **66:**383-394.

18. **Lavigueur, A., H. La Branche, A. R. Kornblihtt, and B. Chabot.** 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. **7:**2405-2417.

19. **Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton.** 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262:**208-214.

20. **Liu, H.-X., M. Zhang, and A. R. Krainer.** 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev. **12:**1998-2012.

21. **Liu, H.-X., and A. R. Krainer.** Unpublished data.

22. **Lynch, K. W., and T. Maniatis.** 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. Genes Dev. **10:**2089-2101.

23. **Mayeda, A., D. M. Helfman, and A. R. Krainer.** 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. **13:**2993-3001. (Erratum, **13:**4458.)

24. **Mayeda, A., and A. R. Krainer.** 1999. Preparation of Hela cell nuclear and cytosolic S100 extracts for in vitro splicing. Methods Mol. Biol. **118:** 309-314.

25. **Mayeda, A., and A. R. Krainer.** 1999. Mammalian in vitro splicing assays. Methods Mol. Biol. **118:** 315-322.

26. **Mayeda, A., and A. R. Krainer.** 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell **68:**365-375.

27. **Mayeda, A., G. R. Screaton, S. D. Chandler, X.-D. Fu, and A. R. Krainer.** 1999. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol. Cell. Biol. **19:**1853-1863.

28. **Ramchatesingh, J., A. M. Zahler, K. M. Neugebauer, M. B. Roth, and T. A. Cooper.** 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol. Cell. Biol. **15:**4898-4907.

29. **Schaal, T. D., and T. Maniatis.** 1999. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol. Cell. Biol. **19:**261-273.

30. **Schaal, T. D., and T. Maniatis.** 1999. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol. Cell. Biol. **19:**1705-1719.

31. **Screaton, G. R., J. F. Cáceres, A. Mayeda, M. V. Bell, M. Plebanski, D. G. Jackson, J. I. Bell, and A. R. Krainer.** 1995. Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. EMBO J. **14:**4336-4349.

32. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. Nucleic Acids Res. **6:**3387-3398.

33. **Staknis, D., and R. Reed.** 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol. **14:**7670-7682.

34. **Stormo, G. D.** 1990. Consensus patterns in DNA. Methods Enzymol. **183:**211-237.

35. **Sun, Q., A. Mayeda, R. K. Hampson, A. R. Krainer, and F. M. Rottman.** 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev. **7:**2598-2608.

36. **Tacke, R., and J. L. Manley.** 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. EMBO J. **14:**3540-3551.

37. **Tian, H., and R. Kole.** 1995. Selection of novel exon recognition elements from a pool of random sequences. Mol. Cell. Biol. **15:**6291-6298.

38. **Tian, M., and T. Maniatis.** 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell **74:**105-114.

39. **Tian, M., and T. Maniatis.** 1994. A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev. **8:**1703-1712.

40. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7:**407-418.

41. **Watakabe, A., K. Inoue, H. Sakamoto, and Y. Shimura.** 1989. A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin μ chain pre-mRNAs. Nucleic Acids Res. **17:**8159-8169.

42. **Wu, J. Y., and T. Maniatis.** 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75:**1061-1070.

43. **Xu, R., J. Teng, and T. A. Cooper.** 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. **13:**3660-3674.

44. **Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth.** 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. Science **260:**219-222.

45. **Zhuang, Y. A., A. M. Goldstein, and A. M. Weiner.** 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl. Acad. Sci. USA **86:**2752-2756.

46. **Zuo, P., and T. Maniatis.** 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev. **10:**1356-1368.

FIGURE LEGENDS

FIG. 1. Experimental procedure for functional SELEX. The structure of the IgM M1-M2 minigene pre-mRNA is shown. The characterized natural ESE (71 nt) was replaced by 20 nt of randomized sequence using overlap-extension PCR (20). A T7 promoter (black box) was built into the PCR product. In-vitro-transcribed RNA was then incubated under splicing conditions. Any spliced mRNA molecules must contain a functional ESE or "winner sequence", designated by the white W in a black box. The spliced mRNA molecules were purified from a denaturing polyacrylamide gel, re-amplified by RT-PCR and cloned. Individual clones were sequenced and analyzed by a sampling algorithm to define a common motif, and a subset were rebuilt into minigene templates, transcribed and assayed for splicing in vitro.

FIG. 2. Splicing of the initial RNA pool. 20 fmol of wild-type IgM minigene pre-mRNA (W; lane 1), ESE-deleted pre-mRNA (ED; lane 2), or a pre-mRNA pool representing 1.2 x 10^{10} different molecules with a randomized 20-nt segment within exon M2 were spliced in 25-μl reactions in nuclear extract (lane 3), S100 extract alone (lane 4) or S100 extract complemented by 20 pmol of recombinant SC35 (lane 5). The structures of the precursor, intermediates and products are indicated next to the autoradiogram. The expected size of the spliced mRNA product of the ESE deletion mutant is indicated by an arrow.

FIG. 3. Analysis of the SC35-selected sequences. (A) Sequence alignment and identification of a consensus motif. The consensus motif and score matrix were derived as described (20). The sequences were aligned on the basis of the highest score motif for each sequence. Nucleotides matching the consensus are shown white on black; mismatched nucleotides are not shaded. The scores of the aligned motifs are indicated on the right. Additional motifs present in some of the sequences with a score greater than

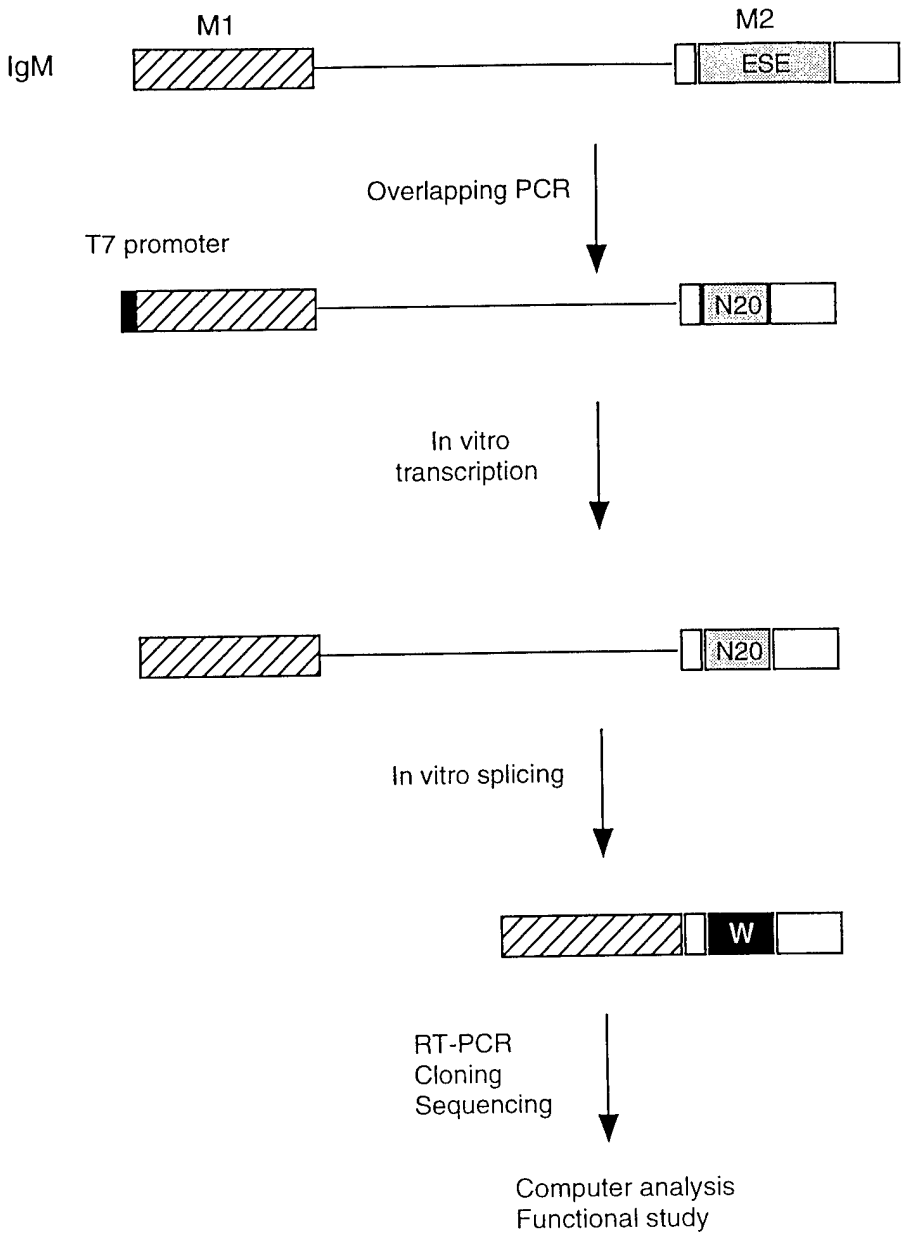
1.62 (the mean score of the random pool) are underlined. In two cases these include a trinucleotide contributed by the 3'-flanking sequence, which is indicated with lowercase letters (cua). The consensus shown is only an approximation that indicates the most frequent nucleotide(s) at each position. The lowercase c at position 5 denotes a slight preference for this nucleotide over the other three nucleotides, which occur at similar frequencies. Y = pyrimidine; S = G or C; R = purine. The nucleotide composition of the selected pool is shown at the bottom. The nucleotide composition of the initial RNA pool was: A = 21%, G = 39%, C = 19%, U = 21%. (B) Representation of the SC35 ESE score matrix and consensus motif. The diagram shows the frequency of each nucleotide at each position of the octamer consensus, adjusted for the compositional bias of the initial pool (20). The height of each letter is proportional to its frequency, and the nucleotides are shown from top to bottom in decreasing order of frequency. This method of displaying nucleotide frequencies is based on references 3 and 13.

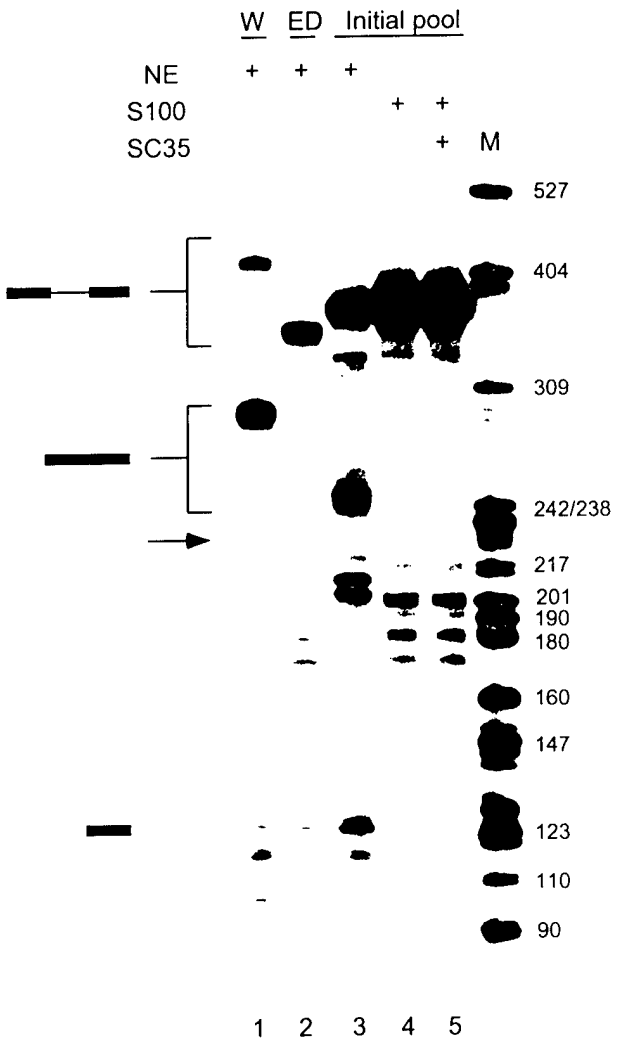
FIG. 4. Activity of the selected ESE motifs. SC35 SELEX winners were rebuilt into the IgM M1-M2 minigene by overlap-extension PCR (as in Fig. 1), and transcripts corresponding to individual winners were spliced in HeLa nuclear extract (lanes 1, 4, 7, 10, and 13), in S100 extract alone (lanes 2, 5, 8, 11, and 14), or in S100 extract complemented by recombinant SC35 (lanes 3, 6, 9, 12 and 15).

FIG. 5. Specificity of the selected ESE motifs. (A) Splicing of SC35-selected ESEs was analyzed in nuclear extract (lanes 1, 5, 9, 13 and 17), S100 extract alone (lanes 2, 6, 10, 14 and 18), or S100 extract plus recombinant SC35 (lanes 3, 7, 11, 15 and 19) or recombinant SF2/ASF (lanes 4, 8, 12, 16 and 20). (B) Splicing of SF2/ASF-selected ESEs was examined in nuclear extract (lanes 1, 5, 9, 13 and 17), S100 extract alone (lanes 2, 6, 10, 14 and 18), or in S100 extract plus SC35 (lanes 3, 7, 11, 15 and 19) or SF2/ASF (lanes 4, 8, 12, 16 and 20).

FIG. 6. Comparison of SC35 winner sequences and SC35-ESE motif in two different exonic contexts. (A) The 20-nt D2 winner sequence (Fig. 3), a variant of D2 with two nucleotide changes to introduce a maximum-score consensus (D2C), and a 19-nt SC35 SELEX winner sequence (6-24) described in a previous study (30) were inserted into the IgM M1-M2 minigene in place of the natural ESE in exon M2, and the corresponding transcripts were spliced in nuclear extract (lanes 2-4). The control pre-mRNA lacking an ESE (ED) is shown in lane 1. (B) The same D2, D2C, and 6-24 sequences were tested in the context of an IgM C3-C4 minigene (27). The Ca pre-mRNA includes the first 38 nt of the C4 exon (lane 1). In the remaining pre-mRNAs, this segment of the C4 exon is followed by the next 38 nt of the C4 exon, which comprise a natural SC35-dependent ESE (lane 2, CaCb), or it is followed by the D2, D2C, or 6-24 sequences (lanes 3-5). The sequences of the relevant portions of the 3' exons are shown below each panel. The two nucleotide changes in D2C compared to D2 are underlined. The mobilities of the pre-mRNAs, mRNAs, and 5' exon intermediates are indicated next to each autoradiogram.

FIG. 7. Correlation between predicted ESE motifs in natural genes and SR-protein-specificity of the pre-mRNAs. Score matrices derived for SC35 and SF2/ASF were used to search the sequences of natural genes. The resulting scores (Y-axis) were plotted against the nucleotide position along each exon (X-axis). The vertical bars indicate the first nucleotide of each motif. SC35 high-score motifs are shown in green and SF2/ASF ones in blue. Since different score matrices were used for each protein, the numerical scores of the two different proteins cannot be compared. (A) High-score motifs in the IgM gene M2 exon. The characterized ESE is indicated by the horizontal magenta bar and the yellow bar indicates the region comprising a recently described silencer (15, 40). (B) High-score motifs in the IgM gene C4 exon. (C) High-score motifs in the Tat gene T3 exon. The horizontal yellow bar indicates the position of the SC35-specific silencer.


M1
M2
IgM
ESE
Overlapping PCR
T7 promoter
N20
In vitro
transcription
N20
In vitro splicing
W
RT-PCR
Cloning
Sequencing
Computer analysis
Functional study

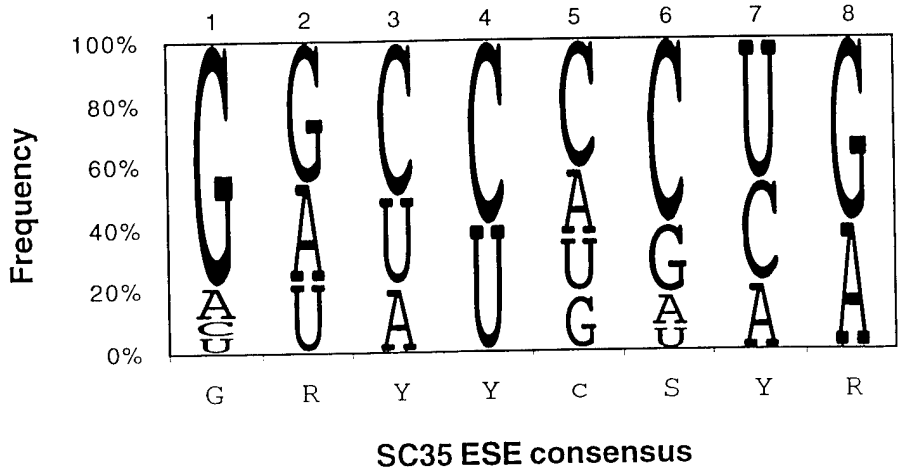

W
ED
Initial pool
NE
S100
SC35
M
527
404
309
242/238
217
201
190
180
160
147
123
110
90
1
2
3
4
5

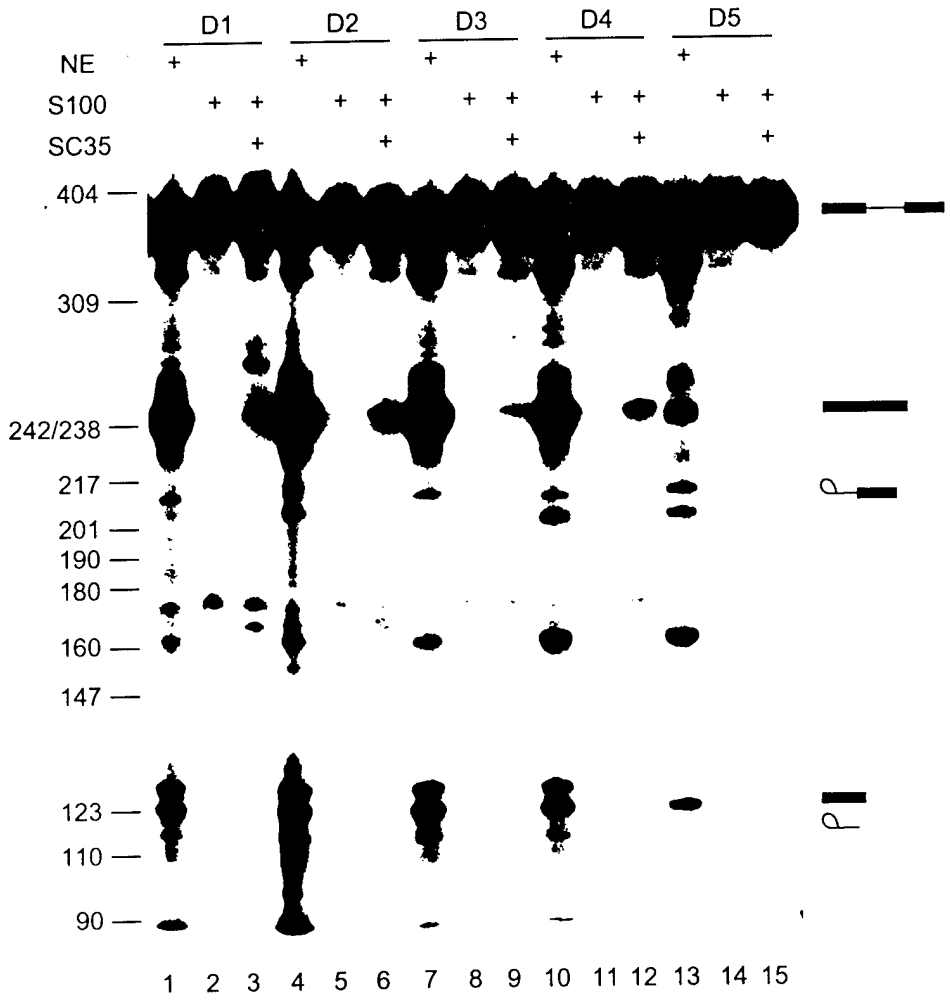
A

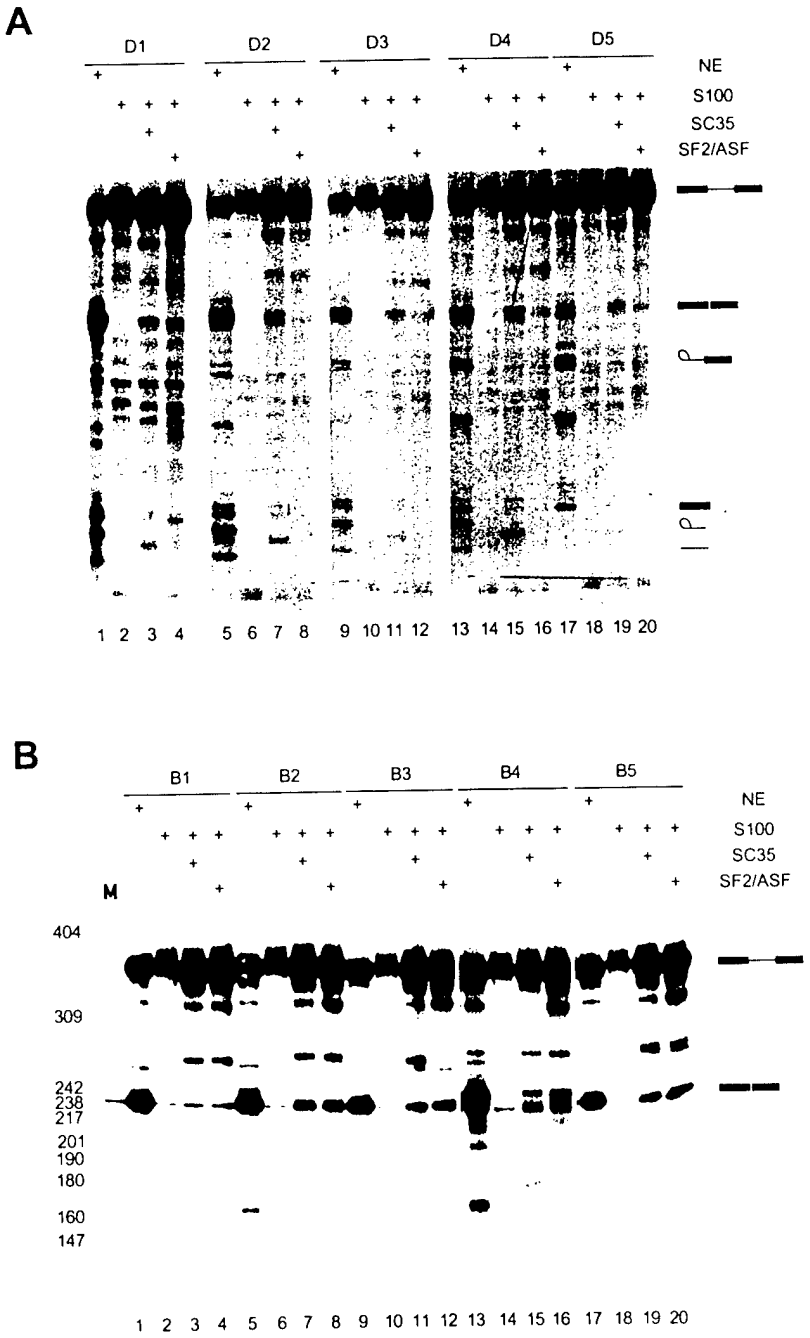
Clone		Sequence	Score
SC35-1		GGAGACUCAAGAUCCCCGG	3.12
SC35-2	(D1)	GACGGCUCGUGCCAUCCUGG	3.55
SC35-3	(D2)	GGGGACAUACUCGGCCGCAG	3.10
SC35-4		GGCGCGGGAUCGCUGGUAUUcua	3.08
SC35-5		CAUGGGGCCCACAAAGAGCGCGG	2.57
SC35-6	(D3)	GGUUGGCGCCAUAUGGUG	2.67
SC35-7		GGAGUCCUCCGUUAUGAUAG	2.84
SC35-8		GUCCUCAGAUGUCCCCUG	3.43
SC35-9		GUUCUGUACUUUUGGGGC	2.24
SC35-10		GAGUGGAAUACCGAGGAAGC	2.40
SC35-11		CCUGGAGACUGGGGACGCUGcua	3.12
SC35-12		GUAAUAGGGAGCGGACCGUA	2.79
SC35-13	(D4)	GUCUAACGGUGCCGGACC	1.88
SC35-14		CAUGCAGCCUCAGAUGGGGA	1.99
SC35-15		ACAGCCGCGCGGGAUGGAGU	2.23
SC35-16		GGACUGUAUUGUUAGGAUGG	2.51
SC35-17		GGAGGUUGGCCCGGUUGUUG	2.41
SC35-18		CGGGAGCACUGUGUUCCGGA	1.56
SC35-19	(D5)	AGUGUUACUACAGCUAUCGC	1.19
SC35-20		GCGGCAGGCUCCAAAGAUGU	3.12
SC35-21		GGAGGUGAUAGGGAUCCGGU	1.96
SC35-22		UGCAAGUGUUGACCUGCGGG	2.70
SC35-23		CCUGGCAUGAACGUUUCGAG	2.31
SC35-24		CGGUCGCCGGGUUAAUGCG	3.06
SC35-25		GAGGGGCGGUCAGUGGCGAC	2.90
SC35-26		GAACUGGCUGAUGGAGUUGC	2.81
SC35-27		GUGGGAAGCGCCCUUGUACA	2.19
SC35-28		GUAAGCUCCCAACGAAGCGG	2.12
SC35-29		AGGCCGACCGGUGUGAUCUG	2.94
SC35-30		CAAGUACGGACUAGAAAACA	2.00
CONSENSUS		GRYYcSYR	MEAN 2.56

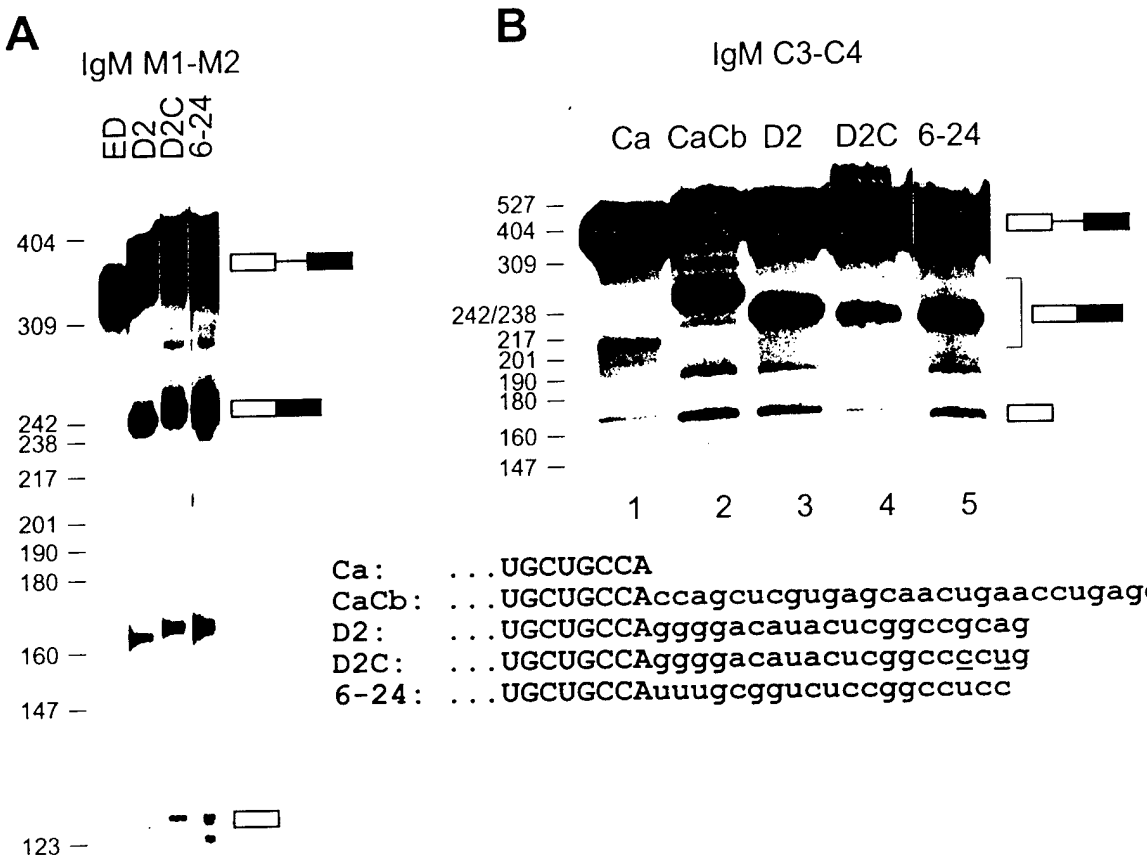
A = 20% G = 37% C = 23% U = 20%

B



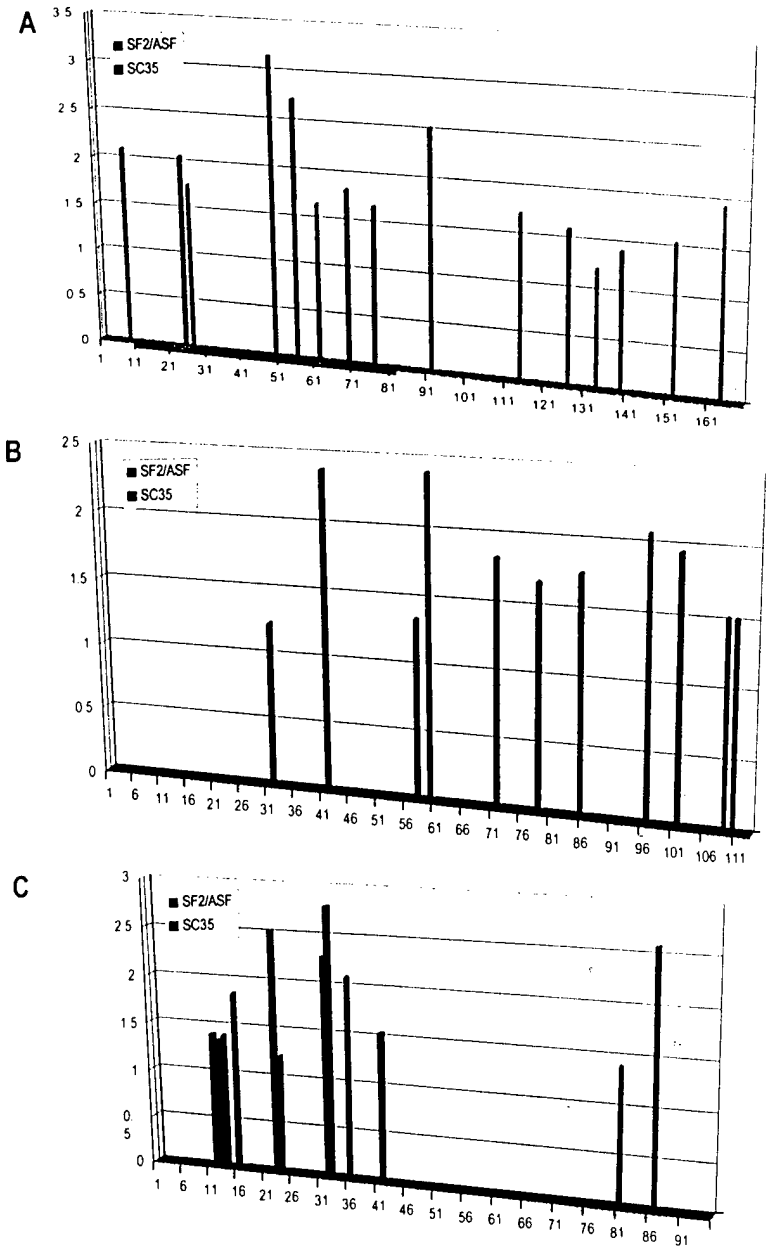

D1
D2
D3
D4
D5
NE
S100
SC35
404
309
242/238
217
201
190
180
160
147
123
110
90
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15


A
D1
D2
D3
D4
D5
NE
S100
SC35
SF2/ASF
P
P
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
B1
B2
B3
B4
B5
NE
S100
SC35
SF2/ASF
M
404
309
242
238
217
201
190
180
160
147
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20




```
Ca:     ...UGCUGCCA
CaCb:   ...UGCUGCCAccagcucgugagcaacugaaccugagggagucagccac
D2:     ...UGCUGCCAggggacauacucggccgcag
D2C:    ...UGCUGCCAggggacauacucggccccug
6-24:   ...UGCUGCCAuuugcggucuccggccucc
```

```
ED:     ...UCUCAGCAUCUAGUAAAC...
D2:     ...UCUCAGCAUggggacauacucggccgcagCUAGUAAAC...
D2C:    ...UCUCAGCAUggggacauacucggccccugCUAGUAAAC...
6-24:   ...UCUCAGCAUuuugcggucuccggccuccCUAGUAAAC...
```


A
SF2/ASF
SC35
3.5
3
2.5
2
1.5
1
0.5
0
1 11 21 31 41 51 61 71 81 91 101 111 121 131 141 151 161
B
SF2/ASF
SC35
2.5
2
1.5
1
0.5
0
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111
C
SF2/ASF
SC35
3
2.5
2
1.5
1
0.5
0
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91

Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins

Hong-Xiang Liu, Michael Zhang, and Adrian R. Krainer

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724-2208 USA

GENES & DEVELOPMENT 12:1998–2012 ISSN 0890-9369/98 $5.00; www.genesdev.org

Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins

Hong-Xiang Liu, Michael Zhang, and Adrian R. Krainer[1]

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724-2208 USA

Using an in vitro randomization and functional selection procedure, we have identified three novel classes of exonic splicing enhancers (ESEs) recognized by human SF2/ASF, SRp40, and SRp55, respectively. These ESEs are functional in splicing and are highly specific. For SF2/ASF and SRp55, in most cases, only the cognate SR protein can efficiently recognize an ESE and activate splicing. In contrast, the SRp40-selected ESEs can function with either SRp40 or SRp55, but not with SF2/ASF. UV cross-linking/competition and immunoprecipitation experiments showed that SR proteins recognize their cognate ESEs in nuclear extract by direct and specific binding. A motif search algorithm was used to derive consensus sequences for ESEs recognized by these SR proteins. Each SR protein yielded a distinct 5- to 7-nucleotide degenerate consensus. These three consensus sequences occur at higher frequencies in exons than in introns and may thus help define exon–intron boundaries. They occur in clusters within regions corresponding to naturally occurring, mapped ESEs. We conclude that a remarkably diverse set of sequences can function as ESEs. The degeneracy of these motifs is consistent with the fact that exonic enhancers evolved within extremely diverse protein coding sequences and are recognized by a small number of SR proteins that bind RNA with limited sequence specificity.

[*Key Words:* SR proteins; exonic splicing enhancers; SF2/ASF; RNA sequence motifs; SELEX]

Received November 26, 1997; revised version accepted April 17, 1998.

Pre-mRNA splicing consists of two *trans*-esterification reactions, which occur in a large RNA–protein complex termed the spliceosome. This high-fidelity process requires precise recognition of the intron–exon borders by the spliceosome. The poorly conserved metazoan splice sites and branch site do not provide sufficient information for this recognition. Additional intron and exon sequences are often necessary for efficient and/or accurate splicing of many higher eukaryotic pre-mRNAs. The positive exon *cis*-acting elements, known as exonic splicing enhancers (ESEs), are often, though not always, found in a purine-rich context. A well-studied example is the ESE in the alternative exon M2 of the mouse *IgM* gene. This 73-nucleotide ESE is essential for splicing of the preceding intron between exons M1 and M2. The M2 ESE can also stimulate splicing of a heterologous regulated intron of the *Drosophila doublesex* (*dsx*) gene. Enhancer activity in the context of the *IgM* pre-mRNA could also be obtained by insertion of certain natural or synthetic purine-rich sequences in place of the natural ESE. However, deletion of the purine-rich sequences within the M2 ESE did not abolish its activity completely (Watakabe et al. 1993; Tanaka et al. 1994). In agreement with this finding, SELEX experiments revealed that certain nonpurine-rich sequences can also function as enhancers (Tian and Kole 1995; Coulter et al. 1997). Most natural ESEs have been identified in tissue-specific or developmentally regulated exons, which typically have weak splice sites and require the ESE for exon inclusion. In some cases, ESEs are specifically recognized by one or more SR proteins (Lavigueur et al. 1993; Sun et al. 1993; Tian and Maniatis 1993, 1994; Ramchatesingh et al. 1995; Gontarek and Derse 1996). In turn, SR proteins are expressed at different levels in different tissues, and their expression also appears to be regulated by alternative splicing (Jumaa et al. 1997; for review, see Cáceres and Krainer 1997).

The SR proteins are a family of highly conserved serine/arginine-rich RNA-binding proteins. They are essential splicing factors (Krainer et al. 1990b, 1991; Ge et al. 1991; Zahler et al. 1992) and also regulate the selection of alternative splice sites in a concentration-dependent manner (Ge and Manley 1990; Krainer et al. 1990a; Zahler et al. 1993a), in part by antagonizing the activity of hnRNP A1 (Mayeda and Krainer 1992). The SR proteins act very early in spliceosome assembly (Krainer et al. 1990a; Fu and Maniatis 1992; Staknis and Reed 1994). They promote the binding of U1 snRNP to the 5′ splice

[1]Corresponding author.
E-MAIL krainer@cshl.org; FAX (516) 367-8453.

site (Eperon et al. 1993; Wu and Maniatis 1993; Kohtz et al. 1994; Staknis and Reed 1994; Zahler and Roth 1995) and of $U2AF^{65}$ to the 3′ splice site (Wu and Maniatis 1993), apparently by interacting with U1 70K and $U2AF^{35}$, respectively. These observations have led to the hypothesis that SR proteins bound to ESEs recruit splicing factors to bind to the splice sites of adjacent introns (Wu and Maniatis 1993; Staknis and Reed 1994).

Nine human SR proteins are presently known: SF2/ASF, SC35, SRp20, SRp40, SRp75, SRp55, 9G8, SRp30c, and the somewhat more divergent p54. These proteins are closely related in primary structure and share the ability to complement splicing in a HeLa cell S100 extract (Ge et al. 1991; Krainer et al. 1991; Fu et al. 1992; Zahler et al. 1992, 1993b; Cavaloc et al. 1994; Screaton et al. 1995; Zhang and Wu 1996). SR proteins appear to have partially redundant functions, such that several different members of the family can complement an S100 extract to splice the same pre-mRNA, and/or stimulate use of the same alternative 5′ splice sites in vitro or in vivo. However, substrate-specific differences in general splicing, enhancer-dependent splicing, or alternative splicing mediated by different SR proteins have also been reported (Fu 1993; Sun et al. 1993; Zahler et al. 1993a; Cáceres et al. 1994; Wang and Manley 1995; Chandler et al. 1997). *Drosophila* SRp55/B52 has been shown to be essential for development (Ring and Lis 1994; Peng and Mount 1995), and at least one copy of the chicken SF2/ASF gene is required for survival of a B-lymphocyte cell line (Wang et al. 1996), demonstrating that at least some functions important for development or cell viability are uniquely carried out by single SR proteins in vivo. Individual SR proteins also differ in their subnuclear localization signals and in their ability to shuttle between the nucleus and the cytoplasm (Cáceres et al. 1997, 1998). Finally, individual SR proteins exhibit striking phylogenetic sequence conservation of all their constituent domains (Birney et al. 1993). Taken together, these observations demonstrate that individual SR proteins have some unique, specific functions.

Although SR proteins have been clearly implicated in ESE recognition and function, predictive rules for the recognition of ESEs by different SR proteins have not been derived. In this study, we sought to determine the specificity of individual SR proteins in ESE recognition by performing a randomization and selection procedure under splicing conditions.

Results

Identification of SR protein target sequences from a random pool under splicing conditions

To find specific target sequences recognized by individual SR proteins under splicing conditions, a procedure based on SELEX (Tuerk and Gold 1990) was utilized imposing a selection for splicing (Tian and Kole 1995; Coulter et al. 1997), rather than for binding (Heinrichs and Baker 1995; Tacke and Manley 1995; Shi et al. 1997; Tacke et al. 1997). We modified the procedure further by carrying out the splicing reactions in the presence of a single, recombinant SR protein, which was used to complement HeLa extracts deficient in SR proteins—either an S100 extract or an SR protein-depleted nuclear extract. We performed the selection for ESEs in the context of a well-characterized *IgM* minigene transcript, comprising the last intron flanked by the M1 and M2 membrane isoform-specific exons. A prototypical ESE was previously mapped to a 73 nucleotide fragment of exon M2 (Watakabe et al. 1993). This ESE was found to be essential for *IgM* pre-mRNA splicing in nuclear or S100 extract (Watakabe et al. 1993; H.-X. Liu et al., unpubl.). The scheme for the randomization and selection procedure is outlined in Figure 1A (for details, see Materials and Methods). First, the natural ESE in the M2 exon (Fig. 1B) was replaced by 20 nucleotides of random sequence. A library of pre-mRNAs representing 1.2×10^{10} different sequences was spliced in S100 extract complemented by either recombinant SF2/ASF, SRp40, or SRp55 (Fig. 2). As a control, equivalent samples from the same library were spliced in nuclear extract, which contains all the SR proteins. A significant proportion of the pre-mRNAs from the initial pool contained sequences that functioned as enhancers in the nuclear extract, thus resulting in easily detectable levels of spliced mRNA (lanes 1,4,7). In contrast, no splicing could be detected by this direct assay in the S100 extract alone (lanes 2,5,8), or in reactions with S100 extract plus one of the SR proteins (lanes 3,6,9). Nevertheless, we assumed that a small proportion of the initial randomized sequences could function as enhancers in the presence of single SR proteins.

The presumptive spliced mRNAs, now carrying functional ESEs, were recovered from the expected region of denaturing polyacrylamide gels. For each S100 complementation reaction, the randomized region of exon M2 of the spliced mRNAs was then amplified by RT-PCR and reassembled into a new pool of pre-mRNAs for further selection. The amplification was carried out by overlap extension with two different upstream primers, to ensure that only spliced mRNAs were amplified (Fig. 1). PCR amplification confirmed the presence of spliced mRNAs after the initial round of selection. Two additional rounds of selection were carried out in SR protein-depleted nuclear extract (Blencowe et al. 1994) complemented with individual SR proteins, to mimic the conditions of nuclear extract, to minimize possible biases specific to the S100 extract, and to select the most efficient ESEs.

After three rounds of selection, the spliced mRNAs were amplified by RT-PCR, subcloned and sequenced. Twenty-four or more independent sequences obtained with each SR protein were analyzed to determine a consensus sequence, using the program Gibbs sampler (Lawrence et al. 1993). The defined motifs were used to generate a score matrix, according to the frequency of each nucleotide at each position. These score matrices were used to search the high-score motifs in each winner sequence. Small portions of the constant flanking regions (18 nucleotides of the 5′ region and 20 nucleotides

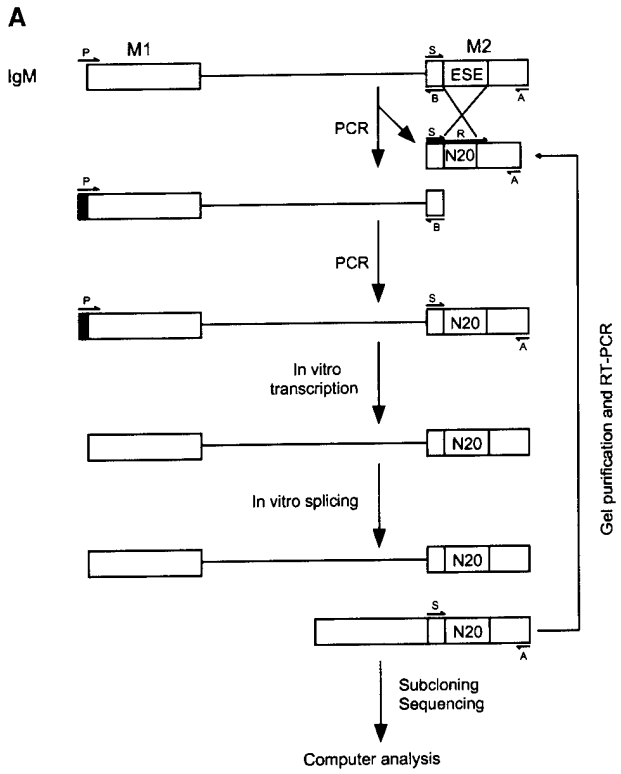


Figure 1. Procedure for randomization and selection of ESEs. (*A*) The natural ESE in mouse *IgM* exon M2 was replaced by a 20-nucleotide segment of random sequence, and a library of pre-mRNAs was constructed by overlap–extension PCR and in vitro transcription. A sample of this pool, representing ~1.2×10^{10} pre-mRNA molecules was then spliced in vitro by complementation of an S100 extract with individual recombinant SR proteins. The pool of spliced mRNA products was gel purified, and the sequences corresponding to the ESE region were rebuilt into pre-mRNA template molecules for a new round of selection, or subcloned and sequenced. The sequences were analyzed by a motif-search algorithm to identify common patterns. (*B*) Sequence of the M2 exon of the mouse *IgM* gene. The sequence of the previously mapped ESE is shown in uppercase.

of the 3′ region) were included during the search. The resulting alignments of sequences selected with SF2/ASF, SRp40, or SRp55 are shown in Figures 3A, 4, and 5, respectively. As a control, 30 sequences from the initial random RNA pool are shown in Figure 3B.

The consensus sequences derived for each of the three SR proteins tested differed in both length and sequence. Each of the consensus sequences is relatively degenerate, and not all of the individual selected sequences match the consensus at every position. However, many of the individual sequences have more than one good match to the consensus, allowing for one or two mismatches.

The SF2/ASF winners gave the consensus sequence SRSASGA (S represents G or C, R represents purine), which only in some cases corresponds to a purine-rich motif. The content of U residues in the SF2/ASF winner pool was 16%, which represents a significant reduction from the 21% of U residues found in the initial random pool. This reduction can be accounted for by the absence of U residues from the consensus motif. The content of C residues increased by 4%. The frequency of A and G did not vary significantly upon selection (Fig. 3A,B). SF2/ASF was shown previously to recognize purine-rich sequences in SELEX procedures based on binding; the reported sequences, RGAAGAAC and AGGACRRAGC (Tacke and Manley 1995), are significantly different from, and simpler than, the consensus motif we found. Similar experiments, performed independently in our lab, revealed a different purine-rich consensus sequence, GARGAGC (A. Hanamura, I. Watakabe, and A.R. Krainer, unpubl.). In the present study, only 13 of 28 winners have uninterrupted purine-rich motifs longer than 5 nucleotides, indicating that SF2/ASF can productively recognize a far broader range of sequences. Indeed, the overall purine composition of the SF2/ASF-selected pool did not change significantly from that of the initial random pool.

The consensus for the SRp40-selected sequences is ACDGS (D represents residues other than C; S represents G or C). This consensus is also very different from that previously determined as an optimal RNA-binding site for SRp40, TGGGAGCRGTYRGCTCGY (Tacke et al. 1997). The content of G residues in the SRp40 winner RNA pool decreased from 39% in the initial random pool

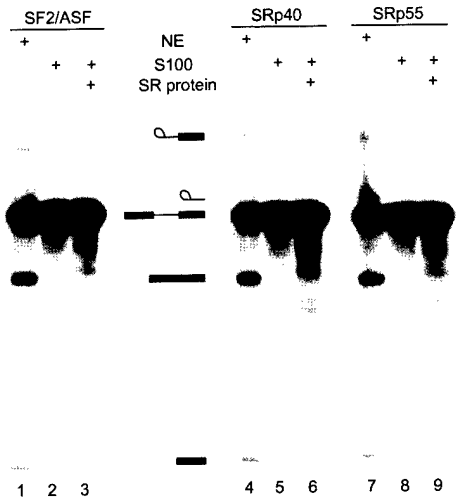


Figure 2. Splicing of the pre-mRNA pool prior to selection. Radiolabeled pre-mRNA (20 fmoles) from the unselected initial pool was incubated under splicing conditions in nuclear extract (lanes *1,4,7*), in S100 extract only (lanes *2,5,8*), or in S100 extract complemented with the indicated SR protein (lanes *3,6,9*). The RNAs were analyzed by denaturing PAGE (12% polyacrylamide) and autoradiography. The structures and mobilities of the precursor, intermediates, and products of splicing (Watakabe et al. 1993) are shown next to each autoradiograph.

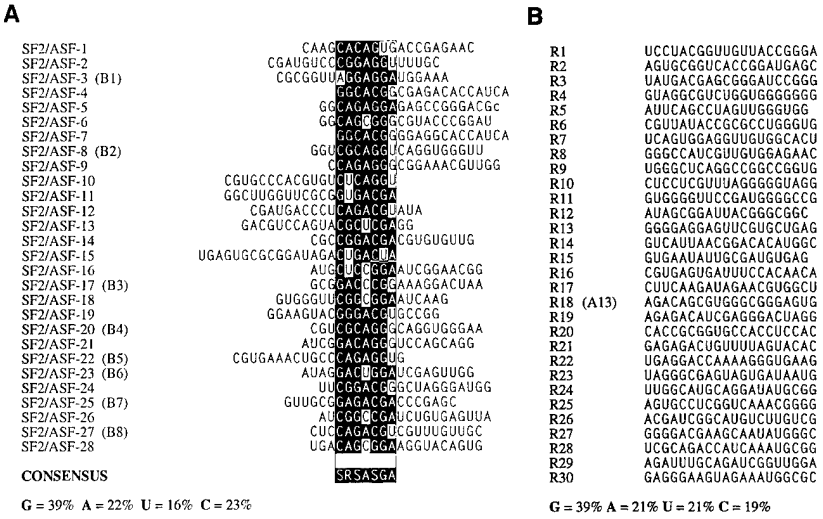


Figure 3. Sequence alignment of SF2/ASF-specific ESEs (*A*) and sequences from the initial random pool (*B*). A consensus motif was identified as described in the text. The sequences are aligned on the basis of the best fit to the consensus within each sequence. Nucleotides in the boxed alignment that match the consensus position are shown white on black; mismatched nucleotides are not shaded. Underlined nucleotides are from the constant region flanking the randomized segment. The nucleotide composition of the randomized segment in the selected pool is provided in the lower left corner. S = G or C; R = A or G. Clones indicated in parentheses were selected for further characterization.

to 34%. The content of C residues increased by 5%. The frequency of A and U did not change significantly (Figs. 4 and 3B). The consensus motif for the SRp40 winners is relatively short but has a sufficient information content, such that, for example, it does not occur by chance in most of the RNAs sampled from the initial random pool. Winners SRp40-1, SRp40-2, SRp40-3, and SRp40-4, for example, are clearly not derived from a single founder sequence by accumulated mutations during PCR. However, they all share the sequence ACGGC, which matches the consensus, and is the only common sequence among these winners. Similar sequence relationships are seen among winners SRp40-5, SRp40-6, SRp40-7; SRp40-8, SRp40-9, SRp40-10; SRp40-11, SRp40-12; SRp40-13, SRp40-14; and SRp40-15, SRp40-16.

The SRp55 winners yielded the consensus sequence USCGKM (S represents G or C; K represents U or G; M represents A or C). The C residue content in the SRp55 winner pool increased significantly, from 19% in the starting pool to 26%. The content of G and U residues decreased by 5% and 4%, respectively (Figs. 5 and 3B). B52, which appears to be the *Drosophila* ortholog of human SRp55, was reported to have GRUCAACCNGGC GACNG as the optimal binding site (Shi et al. 1997). In that report, it was also suggested that a hairpin structure was required for efficient B52 binding. In contrast, we did not observe common secondary structure elements in our human SRp55 winner sequences.

The same sequence analysis programs were used to search the clones from the initial random pool, but no stable pattern was found. Each set of aligned winner sequences was used to create a score matrix that takes into account the overall nucleotide composition of the corresponding winner pool (see Materials and Methods). The scores of the SF2/ASF winner sequences ranged from 1.34 to 3.74, with a mean of 2.7; those of the SRp40 winner pool ranged from 0.33 to 1.68, with a mean of 1.2; those of the SRp55 winner pool ranged from 2.37 to 6.17, with a mean of 4.64. The sequence-scores for each SR protein correlated with the observed splicing efficiencies; however, sequence-scores for different SR proteins cannot be compared. The score matrices were then used

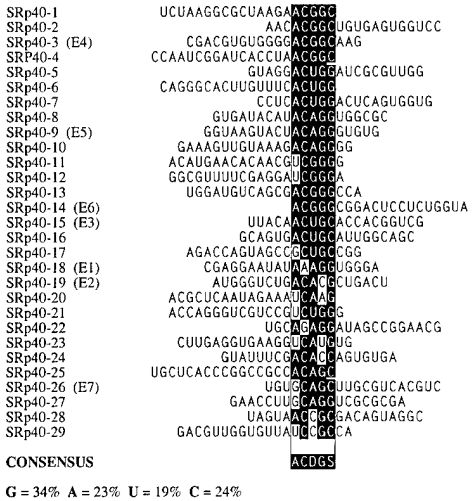


Figure 4. Sequence alignment of SRp40-specific ESEs. D = A, G, or U. For details, see Fig. 3 legend.

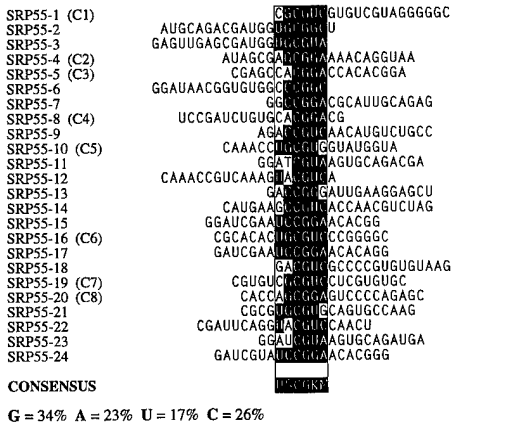


Figure 5. Sequence alignment of SRp55-specific ESEs. K = U or G; M = A or C. For details, see Fig. 3 legend.

to search the clones from all three of the winner pools and the initial random pool. The mean score of the corresponding SR protein-selected winner pool was always higher than that of the other three pools (data not shown). In particular, among the 30 sequences from the initial pool, only 3 had scores above the mean for the SF2/ASF-selected pool, 5 had scores above the mean for the SRp40-selected pool, and none had scores above the mean for the SRp55-selected pool.

The SELEX winner sequences function as bona fide ESEs

To investigate the functional importance of the winner sequences, several winners were randomly chosen from the winner pools of each SR protein. Their ability to function as enhancers was tested by splicing the corresponding pre-mRNAs in HeLa nuclear extract or in S100 extract plus specific SR proteins (Fig. 6).

All the SF2/ASF-selected sequences promoted efficient splicing in nuclear extract (Fig. 6A, lanes 1,4,7,10, 13,16,19,22), compared to an enhancerless construct, which is essentially inactive (Watakabe et al. 1993; data not shown), indicating that the selected sequences could function as true ESEs. In each case, they promoted more efficient splicing in nuclear extract than any of several sampled round 0 sequences (data not shown). Furthermore, these ESE sequences promoted splicing in S100 extract plus recombinant SF2/ASF, albeit with variable efficiencies (Fig. 6A, lanes 3,6,9,12,15,18,21,24), but not in S100 extract only (Fig. 6A, lanes 2,5,8,11,14,17,20,23). The splicing efficiency in S100 extract plus SF2/ASF was lower than that of the nuclear extract. Winner sequences comprising either purine-rich (B1, B3, B4, B5, and B7) or nonpurine-rich motifs (B2, B6, and B8) resulted in a comparable range of splicing efficiencies.

Similar results were obtained in splicing assays with the SRp40 winners. All of the 7 winners tested spliced with somewhat variable efficiencies in nuclear extract (Fig. 6B, lanes 1,4,7,10,13,16,19) and in S100 extract plus recombinant SRp40 (Fig. 6B, lanes 3,6,9,12,15,18,21), but not in S100 extract only (Fig. 6B, lanes 2,5,8,11,14,17,20). The splicing efficiency of the SRp40 winners in nuclear extract was lower on average than that of the SF2/ASF winners (Fig. 6A,B). In S100 extract alone, several of the pre-mRNAs were extensively degraded (Fig. 6B, constructs E1, E2, E4, and E7). This observation is consistent with the notion that SR proteins are involved in the assembly of a commitment complex, such that substrates that are not productively assembled into commitment

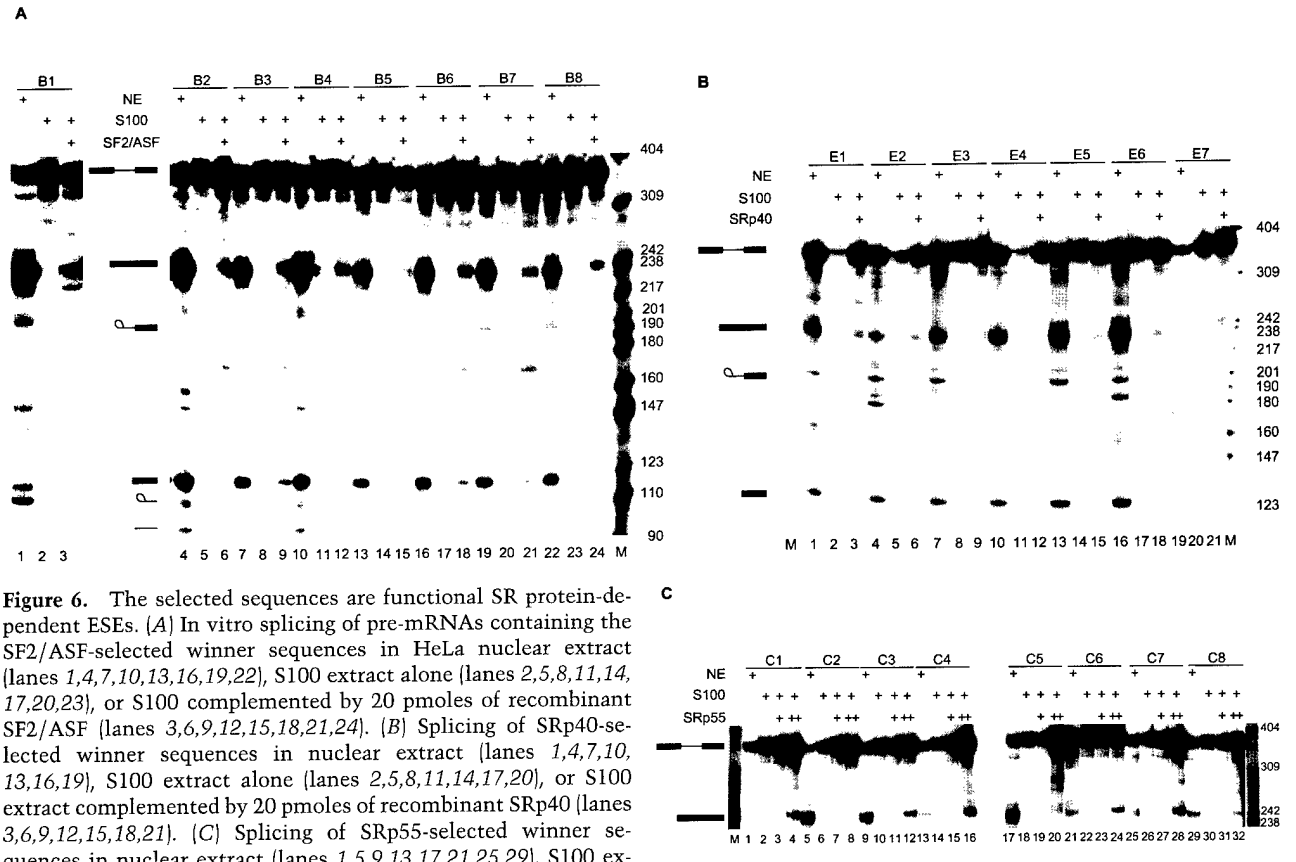


Figure 6. The selected sequences are functional SR protein-dependent ESEs. (*A*) In vitro splicing of pre-mRNAs containing the SF2/ASF-selected winner sequences in HeLa nuclear extract (lanes *1,4,7,10,13,16,19,22*), S100 extract alone (lanes *2,5,8,11,14, 17,20,23*), or S100 complemented by 20 pmoles of recombinant SF2/ASF (lanes *3,6,9,12,15,18,21,24*). (*B*) Splicing of SRp40-selected winner sequences in nuclear extract (lanes *1,4,7,10, 13,16,19*), S100 extract alone (lanes *2,5,8,11,14,17,20*), or S100 extract complemented by 20 pmoles of recombinant SRp40 (lanes *3,6,9,12,15,18,21*). (*C*) Splicing of SRp55-selected winner sequences in nuclear extract (lanes *1,5,9,13,17,21,25,29*), S100 extract alone (lanes *2,6,10,14,18,22,26,30*), or in S100 extract complemented by 10 pmoles SRp55 (lanes *3,7,11,15,19,23,27,31*) or by 20 pmoles of SRp55 (lanes *4,8,12,16,20,24,28,32*). The RNAs were analyzed by denaturing PAGE (5.5% polyacrylamide) and autoradiography. The structures and mobilities of the precursor, intermediates, and products of splicing (Watakabe et al. 1993) are shown next to each autoradiograph.

complexes and pre-spliceosomes are generally more susceptible to degradation by non-specific nucleases present in some batches of extract.

The eight tested SRp55 winners all spliced to variable extents in nuclear extract (Fig. 6C, lanes 1,5,9,13,17,21, 25,29) and in S100 extract plus SRp55 (Fig. 6C, lanes 4,8,12,16,20,24,28,32), but not in S100 extract only (Fig. 6C, lanes 2,6,10,14,18,22,26,30). Interestingly, four of the eight SRp55 winners tested spliced more efficiently in S100 extract plus SRp55 than in nuclear extract alone (Fig. 6C, C1, C4, C6, and C7), suggesting that SRp55 is the only SR protein required for effective recognition of these winner sequences. The higher splicing efficiency of C1, C4, C6, and C7 pre-mRNAs in S100 compared to nuclear extract cannot be accounted for by their increased stability, since the remaining winners, C2, C3, C5, and C7, were also greatly stabilized in S100 extract plus SRp55, but their splicing efficiencies were lower than in the nuclear extract.

We tested whether the short consensus motifs are sufficient to activate splicing. This was done by replacing sequences within template A13 by the short consensus motifs from the B1, B2, C4, or E7 winners (Figs. 3–5) and then testing the splicing activity of the corresponding pre-mRNAs. A13 was isolated from the initial random pool and had very low splicing activity in nuclear extract and no splicing activity in S100 extract complemented with any of the three SR proteins tested. Insertion of one copy of the consensus motifs was sufficient to activate splicing of the modified A13 pre-mRNA in S100 extract plus the cognate SR protein, although the splicing efficiencies were very low (data not shown). The low efficiency suggests that the sequence context surrounding the conserved motifs is also important for splicing, consistent with the observation that several of the winner sequences contain more than one match to the consensus. We also made and tested a number of clustered point mutations in several of the ESEs selected by each SR protein. We were unable to inactivate ESE function either by mutations in the best match to the consensus motif or by mutations on either side of the motif (data not shown). This unexpected observation suggests that each of the selected sequences has a high level of internal functional redundancy, which is probably necessary to allow efficient splicing.

SR protein specificity of the selected ESEs

The fact that each SR protein selected ESEs that fit a different consensus sequence suggests that SR proteins recognize the synthetic ESEs in a sequence-specific manner. On the other hand, the observation that most of the winner sequences promoted more efficient splicing in nuclear extract than in the S100 complementation reactions suggests that SR proteins may function cooperatively. We examined these possibilities by testing the effect of different SR proteins on splicing of pre-mRNAs with each type of winner. These experiments were performed in S100 extract complemented with individual SR proteins or with pairwise combinations thereof. Strikingly, the three kinds of SR protein winner sequences showed very different specificities (Table 1).

The SRp40-selected winner sequences failed to splice in S100 extract plus SF2/ASF (Fig. 7A, lanes 1,4,7,10,13, 16,19), even at higher concentrations of SF2/ASF (data not shown). However, they did splice in S100 extract plus SRp55 (Fig. 7B, lanes 1,4,7,10,13,16). Adding two SR proteins together did not significantly increase the splicing efficiency, although additive effects were observed with two of the winners, E4 and E6, using SRp40 and SRp55 (Fig. 7A,B).

The SF2/ASF-selected winner sequences gave a different result. The B2 winner spliced poorly in the presence of S100 extract and SRp55, whereas the remaining winners, B1, B3, B4, B5, B6, and B7, failed to splice under these conditions (Table 1). SRp40 did not activate splicing of any of the SF2/ASF-selected winners tested. Moreover, SRp40 inhibited splicing of the SF2/ASF-selected winners even in the presence of SF2/ASF (data not shown).

The SRp55 winner C1 spliced in S100 extract plus any of the three SR proteins we examined. However, addition of two SR proteins did not increase its splicing efficiency. All the six other SRp55 winners tested failed to splice in S100 extract plus SF2/ASF or SRp40 (Table 1).

Table 1. *Summary of the activities and specificities of three types of in vitro-selected ESEs*

	No SR proteins	SF2/ASF	SRp40	SRp55
B1	–	+++	±	±
B2	–	++	±	+
B3	–	+++	–	–
B4	–	+++	–	±
B5	–	+	±	±
B6	–	+++	±	±
B7	–	++	±	±
B8	–	+++	N.D.	N.D.
E1	–	–	+++	+++
E2	–	–	++	+++
E3	–	±	++	+
E4	–	–	++	++
E5	–	±	+	++
E6	–	–	+++	+++
E7	–	–	++	++
C1	–	++	++	+++
C2	–	–	–	+
C3	–	±	–	+++
C4	–	–	–	+++
C5	–	–	–	+
C6	–	±	–	++
C7	–	±	±	+++
C8	–	N.D.	N.D.	++

The in vitro-selected ESEs were tested for function as part of *IgM* minigene pre-mRNAs in HeLa S100 extract alone, or in S100 extract complemented with recombinant SF2/ASF, SRp40, or SRp55. The sequences of the B, E, and C winner series are given in Figs. 3A, 4, and 5. (N.D.) Not determined.

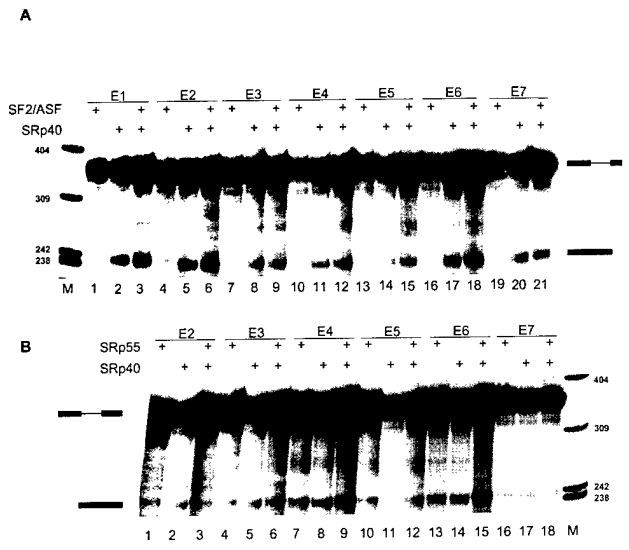


Figure 7. SR protein specificity of in vitro-selected ESEs. (*A*) SRp40-selected ESEs are inactive with SF2/ASF. Splicing was carried out in HeLa S100 extract complemented with 20 pmoles of SF2/ASF (lanes *1,4,7,10,13,16,19*), 20 pmoles of SRp40 (lanes *2,5,8,11,14,17,20*), or 20 pmoles of SF2/ASF plus 20 pmoles of SRp40 (lanes *3,6,9,12,15,18,21*). (*B*) SRp40-selected ESEs can function in the presence of SRp55. Splicing was carried out in S100 extract complemented with 20 pmoles of SRp55 (lanes *1,4,7,10,13,16*), 20 pmoles of SRp40 (lanes *2,5,8,11,14,17*), or 20 pmoles of SRp55 plus 20 pmoles of SRp40 (lanes *3,6,9,12,15,18*).

SR proteins bind specifically to the ESEs in nuclear extract

We have selected ESEs that respond specifically to individual SR proteins under splicing conditions. Because the selection was on the basis of function, it does not necessarily follow that the SR proteins bind directly to the cognate ESEs, although this is generally thought to be the case for at least some natural enhancers (see Discussion). To determine whether SR proteins directly contact the novel ESEs, we carried out UV cross-linking experiments under splicing conditions in nuclear extract. We used radiolabeled RNA fragments comprising the M2 exon with the different ESEs. Twenty femtomoles of an M2 exon RNA comprising the SF2/ASF-selected winner B1 (referred to as B1E) was incubated in the presence of excess cold exon M2 RNA competitors with different ESEs. The reaction mixtures were then irradiated with UV light on ice, digested with RNases A and T1 and analyzed by SDS-PAGE. B1E RNA cross-linked primarily to 34- and 50-kD polypeptides (Fig. 8A, lane 1). The 50-kD product is a nonspecific RNA-binding protein that cross-links to a wide variety of RNAs (data not shown), and is useful as an internal control for recovery and loading. Addition of cold RNA competitors showed that the 34-kD polypeptide bound to the B1E RNA with the greatest specificity (lanes 2,3). Other competitors, comprising an SRp55-selected ESE (lanes 4,5), an SRp40-selected ESE (lanes 10,11), ESEs selected by other SR proteins (lanes 8,9,12,13), or a sequence from the initial random pool (lanes 6,7) failed to compete specifically with B1E for binding to the 34-kD polypeptide (compare the reduction in intensity of the 34-kD band relative to that of the 50-kD band). A minor cross-linked product of 60 kD also decreased in intensity in the presence of the B1E competitor (lane 3), but not in the presence of the other competitors. This may represent an additional protein that interacts specifically with this ESE, or multimerization of the 34-kD product through additional RNA–protein or protein–protein cross-linking.

To confirm that the 34-kD polypeptide is SF2/ASF, we carried out immunoprecipitations after UV cross-linking and RNase digestion (Fig. 8B) (Sun et al. 1993). As expected, the 34-kD polypeptide was the major crosslinked product immunoprecipitated by a monoclonal antibody specific for SF2/ASF (lane 2) but not by a control monoclonal antibody (lane 3).

Similar UV cross-linking experiments were attempted with SRp40- and SRp55-selected ESEs. Cross-linked pro-

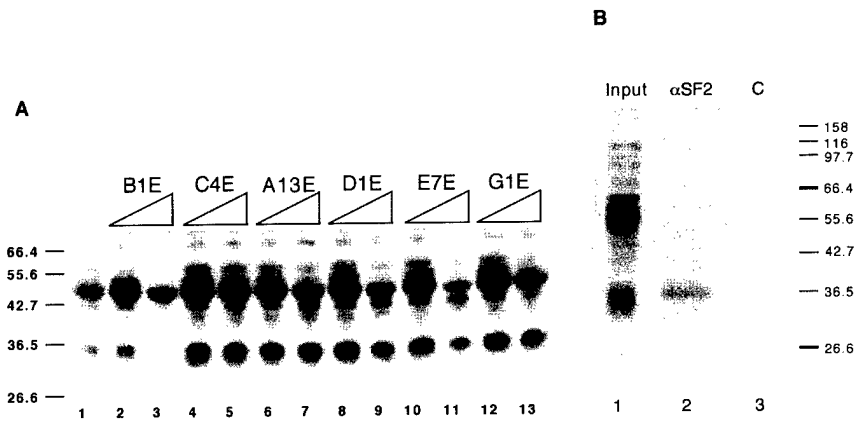


Figure 8. Specific binding of SF2/ASF to an SF2/ASF-selected ESE. (*A*) UV cross-linking competition binding assay. Radiolabeled exon M2 RNA (20 fmoles) comprising the B1 winner sequence (B1E) was incubated under splicing conditions in HeLa nuclear extract. Subsequent UV cross-linking and RNase digestion resulted in label transfer predominantly to two proteins of 34 and 50 kD. The former, which binds specifically, is presumed to be SF2/ASF (see below). Cold competitor RNAs containing either the B1 winner insert, an SRp55-selected insert (C4E), an SRp40-selected insert (E7E), or other control sequence inserts, were present in excess, as indicated above the autoradiograph. (Lane *1*) No competitor; in the remaining lanes, the indicated competitors were present in 5-fold excess (even lanes) or 50-fold excess (odd lanes) over the labeled B1E RNA. (*B*) Immunoprecipitation of SF2/ASF UV cross-linked to the B1E RNA. UV cross-linking was carried out as in *A*, lane *1*. A 5% equivalent of the input was loaded directly (lane *1*). Parallel reactions were incubated with a control antibody (lane *3*), or with anti-SF2/ASF monoclonal antibody (lane *2*), and the immunoprecipitates were recovered in SDS gel loading buffer. In *A* and *B*, the samples were analyzed by SDS-PAGE and autoradiography.

teins of ~40 and 55 kD were detected by use of radiolabeled exon M2 RNAs corresponding to the SRp40 winner E7 (E7E), and the 55-kD protein also cross-linked to the SRp55 winner C4 (C4E), although the background was high (data not shown). Neither of these RNAs cross-linked to proteins with the mobility of SF2/ASF. Cross-linking to the 55-kD protein was competed by an excess of cold C4E RNA, but not by B1E or E7E. Immunoprecipitation with a polyclonal antiserum that recognizes both SRp40 and SRp55 (Du et al. 1997) selectively precipitated cross-linked proteins of the expected size (data not shown). These results suggest that, similar to SF2/ASF, SRp55 and SRp40 also interact directly with their cognate in vitro-selected ESEs.

Sequences that fit the consensus for in vitro-selected ESEs are present in natural exons and known ESEs

Sequences identified by SELEX procedures do not necessarily correspond to functional elements that have evolved in nature (Irvine et al. 1991). To evaluate the biological significance of the novel ESE consensus sequences we identified, we analyzed their distribution in known sequences of natural genes. We reasoned that if the short consensus motifs derived from the in vitro-selected ESEs are akin to natural ESEs, they should be present with higher probability in regions corresponding to known ESEs than elsewhere in the exons or in intron sequences. The score matrices derived for each of the three SR proteins tested were used to search genes or exons with previously characterized ESEs. The resulting scores were then plotted against the position along the exons or genes (Fig. 9).

The natural sequence of the mouse *IgM* exon M2 was analyzed first, since our ESEs were selected in the context of this exon, after deletion of its natural ESE. Remarkably, a high density of motifs with high-score matches to the SF2/ASF and SRp40 consensus was found within the 73-nucleotide natural ESE mapped previously (Watakabe et al. 1993). In contrast, few matching sequences were found in the flanking regions of the exon, and most of these had lower scores, correlating with the lack of splicing upon deletion of the natural ESE. The distribution of motifs with high-score matches to the SRp55 ESE consensus did not correlate with the location of the natural ESE. The SR protein specificity of the natural M2 ESE was not known from previous work, but we have determined that IgM minigene transcripts comprising the natural ESE can function in S100 extract complemented with SF2/ASF, SRp40 or SRp55 (data not shown). The high-score SF2/ASF and SRp40 motifs are present in clusters, suggesting that multiple copies of these motifs are particularly effective as ESEs, or provide an optimal context. Indeed, multimerization of short repeats often results in increased ESE activity, both in natural and synthetic enhancer elements (Tian and Maniatis 1993; Tanaka et al. 1994; Tacke and Manley 1995).

We next analyzed the sequence of the last exon of the bovine growth hormone (bGH) gene, which contains a natural ESE previously mapped to a 115-nucleotide fragment, that is required for splicing of the preceding intron (Sun et al. 1993). The highest density of sequences matching the SF2/ASF, SRp40, and SRp55 consensus ESEs was found within the 115-nucleotide fragment corresponding to the natural ESE, compared to the rest of the 302-nucleotide exon. The highest scores for each of the three SR protein motifs were all found within the fragment with natural enhancer activity. Although the last intron of the bGH pre-mRNA does not splice in S100 extract in the presence of SR proteins, as it apparently requires additional factors, the ESE in the last exon was previously shown to bind SF2/ASF specifically, and this SR protein also stimulated bGH splicing in nuclear extract (Sun et al. 1993).

The *caldesmon* pre-mRNA is alternatively spliced in a tissue-specific manner (Humphrey et al. 1995). An alternative 5′ splice site within the large exon 5 is used in nonmuscle cells, which also exclude exon 6. In smooth muscle, the entire exon 5 is included and spliced to exon 6. A 32-nucleotide repeat present in multiple copies within the 3′ portion of exon 5 functions as an ESE to enhance usage of the upstream non-muscle-specific 5′ splice site (Humphrey et al. 1995). Our sequence analysis showed that SF2/ASF and SRp40 ESE consensus sequences are highly enriched within the 3′ portion of exon 5, whereas SRp55 consensus sequences are found much more frequently upstream of the non-muscle-specific 5′ splice site.

Female-specific alternative splicing of the *Drosophila dsx* pre-mRNA involves six 13-nucleotide repeat elements (dsxRE) and a purine-rich element (PRE) (Tian and Maniatis 1993; Lynch and Maniatis 1995). These *cis*-acting elements are essential for splicing of a *dsx* pre-mRNA in HeLa cell nuclear extract. UV cross-linking analysis showed that the fourth and fifth dsxREs bind specifically to the human SR protein 9G8, whereas the PRE binds preferentially to SF2/ASF and probably other SR proteins of similar size in HeLa nuclear extracts (Lynch and Maniatis 1996). Consistent with these results, our sequence analysis did not reveal any high-score motifs matching the SF2/ASF, SRp40, and SRp55 ESE consensus sequences within the fourth and fifth dsxREs, whereas high-score matches to the SF2/ASF ESE were found within the PRE.

Finally, we also analyzed the sequences of characterized ESEs present in exon 5 of chicken cardiac troponin T (Xu et al. 1993), in exon 3 of the Tat gene of equine infectious anemia virus (Gontarek and Derse 1996), in late pre-mRNAs of bovine papilloma virus type 1 (Zheng et al. 1996), in exon ED-A of human fibronectin (Lavigueur et al. 1993; Caputi et al. 1994), and in the exon downstream of the tat–rev intron of HIV-1 (Amendt et al. 1995; Staffa and Cochrane 1995). In all cases, the sequence analysis was consistent with the available data on these natural ESEs and the binding of SR proteins, when known (data not shown).

Next, we used the same score matrices to analyze the distribution of high-score motifs in human exons versus introns. A total of 570 intron-containing genes, corresponding to 2634 exons (431 kb) and 2079 introns (1300

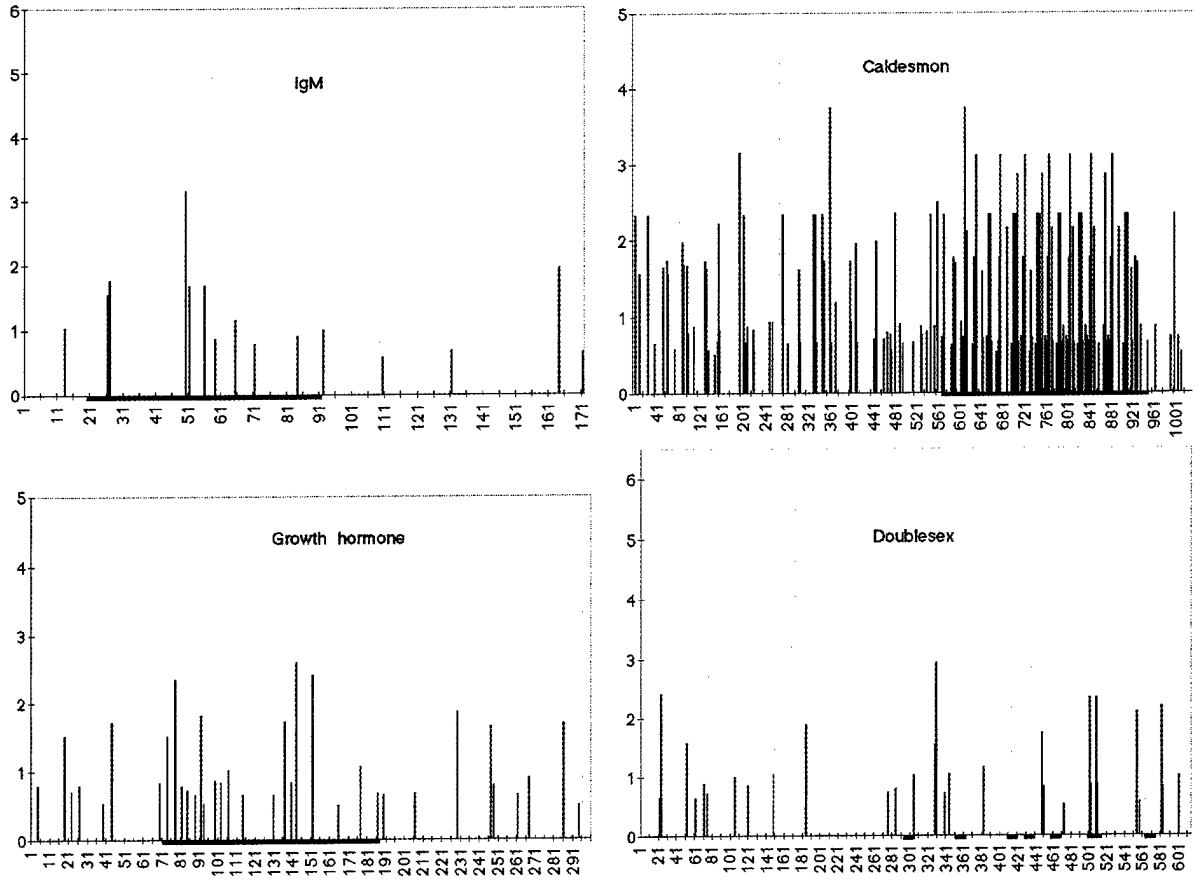


Figure 9. Distribution of in vitro-selected ESE consensus sequences within exons comprising natural ESEs. Score matrices were built for each class of in vitro-selected ESE, as described in Materials and Methods. The indicated natural exon sequences were searched with each score matrix, and the resulting scores (*y*-axis) were plotted against the nucleotide positions for each exon (*x* axis). Note that the *x*-axis scales are different in each case, because of the different exon sizes. Graphs are shown for mouse IgM exon M2, bovine growth hormone 3′ exon, *Drosophila dsx* female-specific exon, and chicken *caldesmon* exon 5. High score motif matches are shown by blue (SF2/ASF), red (SRp40), and yellow (SRp55) vertical bars. For each SR protein, only the sequence matches with a score greater than that of the lowest scoring winner sequence in Figs. 3–5 are shown. Note that there is no relation between the height of bars of different colors. The green horizontal bars under the *x* axis indicate previously mapped ESEs or the *dsx* PRE. The black horizontal bars denote the *dsx* repeat elements (dsxREs).

kb), were extracted from the ALLSEQ data (Burset and Guigo 1996) and analyzed. We searched all sequences with a score equal to or greater than the mean score of the selected winner pool for each SR protein. Remarkably, high-score motifs matching each of the three SR protein ESE consensus sequences were found more frequently in exons than in introns. For SF2/ASF, the density of high-score motifs was 4.3 per kilobase of exon and 2.9 per kilobase of intron; for SRp40, the corresponding numbers were 7.9 per kilobase of exon and 6.8 per kilobase of intron; and for SRp55, they were 5.5 per kilobase of exon and 4.9 per kilobase of intron. The higher density of high-score motifs in exons than in introns is statistically significant because of the large database size, and the *P*-values for these pairwise comparisons were all $<10^{-10}$.

Discussion

We have developed a method to identify ESE elements that can function specifically with individual SR proteins. This goal was accomplished by use of SR protein-deficient HeLa extracts complemented with individual SR proteins, and a pool of pre-mRNAs derived from mouse *IgM*, whose natural ESE in exon M2 was replaced by a 20-nucleotide segment of random sequence. Sequence analysis revealed that the motifs identified by the selection for function with SF2/ASF, SRp40, or SRp55 tend to be clustered in regions corresponding to known, natural ESEs, compared to other exon regions.

The initial randomized pool of *IgM*-derived substrates consisted of 20 fmoles of pre-mRNA ($\sim 1.2 \times 10^{10}$ molecules), which is large enough to include all possible 16-mers ($\sim 4.3 \times 10^{9}$). The longest motif we identified was the 7-nucleotide consensus selected by SF2/ASF, indicating that the initial random pool had sufficient complexity. In parallel SELEX experiments, we also employed a different RNA pool, in which only 14 positions within the *IgM* M2 exon were randomized. Functional

ESEs were also selected out of that library (data not shown), suggesting that the 20-mer library can potentially encode most, if not all, natural ESEs, and that the library size we used was adequate. The functional SELEX procedure was performed for only three rounds. This was deemed sufficient, as all the winner sequences tested proved to be functional. Additional rounds of selection would be expected to result in loss of consensus sequence information, as only the most efficiently spliced RNAs would be recovered.

Our experiments confirm and extend two previous studies that used functional selection from random pools to identify novel ESEs. Tian and Kole (1995) randomized a 20-nucleotide region within the context of a duplicated exon in a model β-globin pre-mRNA. They selected sequences that promoted inclusion of the middle duplicated exon in HeLa nuclear extract. The resulting ESEs after five or seven selection cycles included both purine-rich and nonpurine-rich motifs (Tian and Kole 1995). A related approach was used by Coulter et al. (1997) to identify ESEs that promote inclusion of the internal alternative exon 5 of chicken cardiac troponin T. In this case, the natural ESE was replaced by a 13-nucleotide randomized segment, and the selection for splicing was carried out by three rounds of transient transfection into QT35 quail cells. The resulting ESEs included both purine-rich elements and a novel class of AC-rich elements (ACEs; Coulter et al. 1997). These pioneering studies could not readily identify the factors responsible for ESE recognition—although SR proteins were obvious candidates—because they relied on crude nuclear extracts or cultured cells. In addition, the novel ESEs did not fall into obvious consensus sequences, most likely because they represent a complex collection of elements recognized by several distinct factors. We improved this general approach by performing the selections in extracts dependent upon addition of individual SR proteins, which allowed us to identify functional ESEs recognized and activated by each SR protein, and therefore to derive a corresponding consensus sequence. Our selections were carried out in a different context from those in the above two studies, which focused on inclusion of an optional exon; we used the last exon of the IgM pre-mRNA, which requires an enhancer for splicing of the last intron (Watakabe et al. 1993).

Previous work also attempted to address the specificity of SR proteins in ESE recognition by use of conventional SELEX procedures based on selection for high-affinity binding (Tacke and Manley 1995; Shi et al. 1997; Tacke et al. 1997). These studies gave very different results from our current results with some of the same SR proteins but with selection cycles based on function. First, the consensus sequences obtained by these two approaches are very different. The SF2/ASF motifs defined by binding (Tacke and Manley 1995; A. Hanamura, I. Watakabe, and A.R. Krainer, unpubl.), which are purine-rich, appear to be a subgroup of those defined by function, although they yield relatively low scores when analyzed with our SF2/ASF score matrix. It should be noted that a purine-rich composition is not sufficient for function, but rather, specific sequences are required (Tanaka et al. 1994; Ramchatesingh et al. 1995). Second, many of the winner sequences obtained by binding protocols were not functional as ESEs. In contrast, among the winner sequences obtained by our functional selection protocol, many were tested, and all of these were functional ESEs. Third, the complexity of the winner sequences identified by binding SELEX is much lower than that of the ESEs identified by functional SELEX. As a result, the consensus sequences obtained from the binding selection are less degenerate than those we obtained through functional selection. This may be attributable in part to the use of more selection/amplification cycles in some of the binding SELEX experiments. In the natural situation, exon sequences are obviously very diverse. Degenerate sequence specificity is probably essential for a limited number of SR proteins to be able to recognize a very large number of ESE-containing exons in different genes.

The different results obtained by binding selection and functional selection protocols shed light on the mechanisms of ESE function. The binding selection is based on the affinity of RNA–protein interactions, and the iterative protocol is designed to yield the binding sites with the highest affinity for the protein of interest. However, it appears that the best binding sites are not necessarily the best functional sites, and, in some cases, a high affinity may preclude function. In addition, optimal interactions between an SR protein and its cognate ESEs may require other splicing components, as opposed to just the purified protein. The binding protocol is carried out with the purified protein, whereas the functional selection protocol is carried out in the presence of all components required for splicing. There are also technical reasons why iterative binding protocols may not yield optimal functional sites. The binding affinity and/or the specificity of the binding may be significantly affected by the idiosyncrasies of the binding assay employed (Irvine et al. 1991). Indeed, several of the SF2/ASF and SRp40 winners identified by iterative binding failed to bind to the cognate SR protein when analyzed by a different binding assay (Tacke and Manley 1995; Tacke et al. 1997). Another contributing factor to the discrepancy between the results obtained in binding and functional assays may be that in at least some applications of the former, truncated proteins lacking the RS domain were used (Tacke and Manley 1995). Although the precise functions of the RS domains are not completely understood, they appear to be important for protein–protein and/or RNA–protein interactions (Wu and Maniatis 1993; Tacke et al. 1997; Xiao and Manley 1997). Thus, deletion of the RS domain may affect the binding specificity, as may an incorrect or incomplete phosphorylation state of the domain. In addition, the use of oligo-histidine or other tags in some studies may also affect the binding specificity. In the present study, we used unagged proteins expressed in *Escherichia coli*. Although these proteins are not phosphorylated when isolated, they are very rapidly phosphorylated upon addition to nuclear or S100 extracts (Hanamura et al. 1998). Finally, in the case of the differ-

ent consensus sequences obtained previously for *Drosophila* B52 (Shi et al. 1997) and in the present study for human SRp55, the binding specificity may have diverged considerably between arthropods and vertebrates. For example, the enhancer complex formed on the PRE of the *Drosophila dsx* pre-mRNA binds SRp55/B52 in *Drosophila* Kc cell extracts, but does not appear to bind human SRp55 in HeLa cell extracts (Lynch and Maniatis 1996).

The specific recognition of ESEs by SR proteins is well documented. Some examples include recognition of purine-rich ESEs in bovine growth hormone pre-mRNA by SF2/ASF (Sun et al. 1993), in cardiac troponin T by SRp40 and SRp55 (Ramchatesing et al. 1995), and in the *dsx* pre-mRNA by one of the *Drosophila* SRp30 proteins (Lynch and Maniatis 1996). Our data address the molecular basis of the redundancy and specificity of SR proteins. The different consensus sequences of the three types of in vitro-selected ESEs, and their different responses to individual SR proteins provide an indication of the specificity of SR proteins in ESE recognition and function. The consensus sequence of SF2/ASF-selected ESEs, SRSASGA, matches the sequence of most purine-rich ESEs characterized to date. It is worthwhile to note that this sequence is devoid of U residues. In the *HPRT* and *IgM* genes, the presence of C residues within the purine-rich ESEs was compatible with enhancer function, whereas changing the C residues to U residues abolished their enhancer activity (Tanaka et al. 1994). The SRp55-selected winners also had a reduced U content, and we suspect that a low U composition contributes to the information content that defines ESEs recognized by these SR proteins.

The SRp40-selected ESEs share a relatively short consensus sequence, ACDGS. They could be activated by SRp55, but not by SF2/ASF. The fact that SRp55 could activate SRp40-selected ESEs suggests that these two SR proteins, which are closely related in domain structure, unmodified molecular mass (31.2 kD for SRp40; 39.6 kD for SRp55), and sequence (65% identity; Screaton et al. 1995), also have some functional overlap. It is unlikely that the sequences selected by SRp40 fortuitously comprise a distinct SRp55 recognition site, but not an SF2/ASF site, as all of the seven independent SRp40 ESEs tested had similar properties. When the SF2/ASF score matrix was used to search the seven examined SRp40-selected ESEs, most of them had a score lower than the minimum score of the SF2/ASF-selected ESEs, which could explain why SRp40-selected ESEs were not activated by SF2/ASF. We do not know why E4, which had a score higher than the average score of SF2/ASF-selected ESEs, was not activated by SF2/ASF. However, it is likely that the sequence context, for example, in the form of negative elements or silencers, somehow prevents activation of this motif by SF2/ASF. A related observation is the fact that SRp40 inhibited splicing of some SF2/ASF-selected ESEs even in the presence of SF2/ASF. This may be attributable to formation of inhibitory complexes with SRp40, such that SR proteins may also participate in exonic silencer function, depending on the sequence context. An interesting implication is that the variable expression levels of these antagonistic SR proteins may determine the cell type-specific function of certain ESEs.

We did not observe any cooperative effects among the SR proteins tested. However, the fact that most of the ESEs we identified gave higher splicing efficiencies in nuclear extract than in S100 extract complemented with SR proteins suggests that other SR proteins and/or additional splicing factors may be required for optimal ESE recognition or function. With other substrates, such as β-globin or certain natural ESE-dependent pre-mRNAs, comparable splicing efficiencies can be obtained in the two systems. A few natural or synthetic ESE-dependent pre-mRNAs can only splice in the nuclear extract, apparently because they require one or more unknown factors, distinct from SR proteins, that are also absent in the S100 extract (Sun et al. 1993; Tacke and Manley 1995; Tacke et al. 1997). If this is a property of the ESE, rather than its context, this class of ESEs may not be well represented in the consensus sequences we derived. However, we did find high-score motif matches within one such natural ESE, that of the bGH last exon (Fig. 9). The ESEs we obtained were selected to function in S100 extract plus an SR protein, and hence, it is not surprising that at least basal function could be observed in this complementation system. However, maximal activity appears to require one or more additional factors that may be limiting in the S100 extract.

Many natural ESEs have been found in the last several years. Most of these well-defined ESEs are purine rich, although this nucleotide composition may reflect an experimental bias. First, many of the biochemical studies were carried out in HeLa cell nuclear extract, in which SF2/ASF, which prefers purine-rich sequences, may be the most abundant SR protein. Second, purine-rich motifs may be easier to find by visual inspection which, together with the precedent of known purine-rich ESEs, makes them more likely to be studied further. We have identified three new degenerate motifs, which are not necessarily purine rich. Significantly, the SF2/ASF and SRp40 motifs we defined occur more frequently (and with higher scores) within exon segments corresponding to known ESEs than elsewhere in the exons. All of the motifs also occur more often in exons than in introns and may thus contribute to defining exon–intron boundaries. These consensus sequences may be useful for the prediction of natural ESEs in uncharacterized exons. Our data also suggest that target sites for multiple SR proteins are clustered within natural ESEs. This may explain why large deletions are often required to inactivate natural ESEs. The SRp55 ESE consensus motif did not always correlate with the location of natural ESEs. Interestingly, in the *caldesmon* gene, SF2/ASF and SRp40 sites are enriched in the 3′ portion of exon 5 that is included in smooth muscle cells, whereas SRp55 sites occur more frequently in the 5′ portion of exon 5 that is included in all cell types. Inclusion of the constitutively spliced upstream segment of exon 5 may also require a functional ESE, which we would predict, on the basis of

the present data, to be SRp55-dependent. The differential recognition of the alternative 5′ splice sites associated with the *caldesmon* exon 5 by different SR proteins may be responsible for the proper developmental and tissue-specific expression of *caldesmon* by alternative splicing.

Our results with just three of the nine known SR proteins indicate that a highly diverse set of sequences can function as SR protein-specific ESEs. In fact, of the more than 10^{10} 20-mers we tested in the context of the *IgM* exon M2, a significant proportion contained functional enhancers. Thus, of the RNA molecules from the unselected library that remained at the completion of a splicing reaction in nuclear extract, ~20% were spliced (Fig. 2). Likewise, in the present study and in two previous studies, about 10%–20% of random 20-mers or 13-mers comprised sequences that could enhance splicing significantly above background (Tian and Kole 1995; Coulter et al. 1997; and data not shown). This level of degeneracy in sequence-specific RNA recognition may be essential to allow evolution of effective ESEs interspersed with open reading frames that are constrained by the structure and function of the encoded proteins. These findings imply that many exons are likely to have multiple ESEs that are to some extent redundant, but which may also function additively or allow fine tuning of tissue-specific or developmentally regulated expression. On the other hand, we suspect that splicing silencer elements, which are less well understood, will prove to have similar complexity. As a result, it is likely that many natural sequences that match the simple motifs identified in this study will fail to function as ESEs unless they are placed in an appropriate context.

Materials and methods

Preparation of HeLa cell extracts and recombinant SR proteins

Nuclear and cytosolic S100 extracts were prepared from fresh 12-liter suspension cultures of HeLa cells, as described (Mayeda and Krainer 1998a).

Expression and purification of the authentic form of the recombinant SR proteins SF2/ASF, SRp40 and SRp55, by use of the expression vector pET9c (Novagen), were carried out as described previously (Krainer et al. 1991; Screaton et al. 1995). The integrity and purity of these recombinant SR proteins were checked by SDS-PAGE and their specific activities were determined by in vitro splicing of β-globin pre-mRNA in S100 extract (data not shown).

Randomization and selection

The SELEX procedure is outlined in Figure 1A. The sequence of the wild-type *IgM* exon M2 is shown in Figure 1B. The plasmids μM1-2 and μMΔ, which bear a mouse *IgM* minigene with or without the natural enhancer, respectively (Watakabe et al. 1993), were a generous gift from Prof. Y. Shimura. The randomized substrate pool was constructed by overlap-extension PCR (Horton et al. 1989; Tian and Kole 1995). Two sets of PCRs were performed with μMΔ as template. The first PCR was carried out with primers R and A. The second PCR used primers P and B. The products from the two reactions were then combined and further amplified with primers P and A. The resulting PCR product was then used for in vitro transcription with SP6 RNA polymerase to generate a radiolabeled pre-mRNA substrate pool. The pool of spliced mRNAs generated by in vitro splicing was excised from a urea–polyacrylamide gel, eluted in 0.5 M ammonium acetate plus 0.1% SDS, reverse transcribed by use of Superscript II RT (GIBCO-BRL), and amplified by PCR with primers P and A. The amplified product was further amplified with primers S and A. The PCR product was purified on a 2% agarose gel and reassembled into the pre-mRNA template by overlap-extension PCR for the next round of selection. The reverse transcription and PCR reactions were performed as suggested by the vendors (GIBCO-BRL and Stratagene, respectively). All of the PCRs were carried out with the high-fidelity Pfu DNA polymerase. The primers were purchased from Operon and were used at a concentration of 1 μM. After three rounds of selection, the amplified spliced products were subcloned into the vector PCR-Blunt (Stratagene) and sequenced by use of a Dye Terminator Cycle Sequencing kit (Perkin-Elmer) and an automated ABI 377 sequencer. The second and third rounds of SELEX were performed in nuclear extract depleted of total SR proteins by Mg^{2+} precipitation (Blencowe et al. 1994). The sequences of the primers were as follows: Primer P, 5′-ATTTAGGTGACACTATAGAATAC-3′; primer A, 5′-GCAGGTCGACTCTAGAAAGAAG-3′; primer S, 5′-GTGAAATGACTCTCAGCAT-3′; primer B, 5′-ATGCTGAGAGTCATTTCAC-3′; primer R, 5′-GTGAAATGACTCTCAGCAT(N)$_{20}$-CTAGTAAACTTATTCTTACGTC-3′.

Identification of consensus motifs among the selected sequences

Functional selected sequences for each SR protein were aligned by use of Gibbs sampler (Lawrence et al. 1993), with the assumption that there is a common sequence motif of length L present at least once in all of the sequences. Because Gibbs sampler is a stochastic algorithm, for each fixed L, at least 10 different runs (with different random seeds) were carried out for times sufficient to achieve convergence. A conservative value for L was determined empirically by observing a sharp drop in information per parameter (Lawrence et al. 1993) as L was increased. To exclude the possibility that the predicted consensus motif arose by chance, the information per parameter was also compared to alignments of random sequences obtained by shuffling the nucleotides within each sequence. The final alignment was manually adjusted in a few cases, when better matches to the consensus could be obtained by including a few flanking nucleotides in the alignment.

Construction of a scoring matrix

First, a frequency matrix f_i (a) was calculated from the alignment (i is the position of nucleotide a). Given a background frequency for the set of sequences, $p(a)$, the scoring matrix is defined by the following formula:

$$s_i(a) = \log_2 \frac{f_i(a) + \epsilon p(a)}{p(a)(1 + \epsilon)}$$

where i = (1, 2, ..., L), a = (A, C, G, U), and $\epsilon = 0.5$ is the Bayesian prior parameter (Lawrence et al. 1993).

A motif score is equal to the sum of the scores at each position. Motifs may be ranked by their scores. The top three scores in each sequence using all three different scoring matrices (SF2/ASF, SRp40, and SRp55) were calculated and tabulated (data not shown).

The sequence scores were consistent semiquantitatively with the gel intensity data when the sequence-scores for a given SR protein were defined as follows: (1) The maximum score for each selected sequence was calculated using the scoring matrix, and the threshold was defined as the minimum of these scores; and (2) The sequence score was defined as the number of nonoverlapping motifs that have a score greater than or equal to the threshold. This integer score correlated well with the corresponding gel intensity.

It should be noted that a motif scoring matrix may depend on the pre-mRNA substrate and on the experimental parameters, such as the concentration of SR protein.

In vitro splicing

In vitro splicing was performed as described previously (Mayeda and Krainer 1998b). Briefly, 20 fmoles of ^{32}P-labeled, $^{7}CH_3GpppG$-capped SP6 or T7 transcripts generated from PCR products were incubated in 25-µl splicing reactions. The reactions contained 4 µl of HeLa nuclear extract or 7 µl of S100 extract in buffer D. The $MgCl_2$ concentration was 4.8 mM. Twenty picomoles of the appropriate SR protein was used in S100 complementation assays. After incubation at 30°C for 4 hr, the RNA was extracted and analyzed on 5.5% or 12% polyacrylamide denaturing gels, followed by autoradiography.

UV cross-linking

UV cross-linking experiments were carried out under splicing conditions with or without a 5- to 50-fold molar excess of unlabeled RNA competitor. Polyvinyl alcohol was omitted from the cross-linking reactions. After a 30-min incubation at 30°C the reactions were exposed to 254 nm UV light by use of a Spectronics XL-1000 UV cross-linker at a setting of 1.8 J/cm^2 on ice. RNase A (10 µg) and RNase T1 (100 units) were added and the reactions were incubated for 15 min at 37°C. The cross-linked proteins were analyzed by SDS-PAGE on a 12% gel, followed by autoradiography.

Immunoprecipitations

Immunoprecipitations were performed as described (Sun et al. 1993). The anti-SF2/ASF monoclonal antibody recognizes the amino terminus of SF2/ASF and does not cross-react with other human SR proteins (Hanamura et al. 1998). Polyclonal antiserum against SRp40 (anti-HRS/SRp40) was a generous gift from Drs. K. Du and R. Taub (Du et al. 1997). The cross-linking reactions were precleared after incubation with control antibodies and 50 µl of protein A–agarose (1:1 suspension) in 500 µl of IP buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 0.05% NP-40) for 2 hr at 4°C. An unrelated monoclonal antibody of the same isotype was used for the SF2/ASF preclearing step and rabbit preimmune serum was used for the SRp40 preclearing step. After spinning in a microcentrifuge for 30 min at 4°C, the supernatants were transferred to tubes containing the appropriate antibody immobilized on protein A–agarose and rocked overnight at 4°C. The bound material was recovered by centrifugation, washed twice with 1 ml of IP buffer, eluted in 30 µl of sample buffer [62.5 mM Tris-HCl at pH 6.8, 2% (wt/vol) SDS, 10% (vol/vol) glycerol, 5% (vol/vol) 2-mercaptoethanol], and analyzed by SDS-PAGE and autoradiography.

Acknowledgments

We thank Dr. Y. Shimura for the gift of plasmids µM1-2 and µMΔ, Dr. A. Mayeda for recombinant SR proteins, and other members of our laboratory for sharing reagents and valuable ideas. We are grateful to Drs. A. Mayeda, M. Murray, L. Cartegni, and D. Horowitz for comments on the manuscript. This work was supported by the National Institutes of Health (NIH) grants GM42699 to A.R.K. and HG00010 to M.Z., and by fellowships from the Cold Spring Harbor Laboratory Association and from the U.S. Army Medical Research and Materiel Command under DAMD 17-96-1-6172 to H.-X.L.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Amendt, B.A., Z.-H. Si, and C.M. Stoltzfus. 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: Evidence for inhibition mediated by cellular factors. *Mol. Cell. Biol.* **15:** 4606–4615.

Birney, E., S. Kumar, and A.R. Krainer. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21:** 5803–5816.

Blencowe, B.J., J.A. Nickerson, R. Issner, S. Pennman, and P.A. Sharp. 1994. Association of nuclear antigens with exon-containing splicing complexes. *J. Cell Biol.* **127:** 593–607.

Burset, M. and R. Guigo. 1996. Evaluation of gene structure prediction programs. *Genomics* **34:** 353–367.

Cáceres, J.F. and A.R. Krainer. 1997. Mammalian pre-mRNA splicing factors. In *Eukaryotic mRNA processing* (ed. A.R. Krainer), pp. 174–212. IRL Press, Oxford, UK.

Cáceres, J.F., S. Stamm, D.M. Helfman, and A.R. Krainer. 1994. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* **265:** 1706–1709.

Cáceres, J.F., T. Misteli, G.R. Screaton, D.L. Spector, and A.R. Krainer. 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* **138:** 225–238.

Cáceres, J.F., G.R. Screaton, and A.R. Krainer. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & Dev.* **12:** 55–66.

Caputi, M., G. Casari, S. Guenzi, R. Tagliabue, R. Sidoli, C.A. Melo, and F.E. Baralle. 1994. A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res.* **22:** 1018–1022.

Cavaloc, Y., M. Popielarz, J.P. Fuchs, R. Gattoni, and J. Stévenin. 1994. Characterization and cloning of the human splicing factor 9G8: A novel 35 kD factor of the serine/arginine protein family. *EMBO J.* **13:** 2639–2649.

Chandler, S.D., A. Mayeda, J.M. Yeakley, A.R. Krainer, and X.-D. Fu. 1997. RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc. Natl. Acad. Sci.* **94:** 3596–3601.

Coulter, L., M. Landree, and T. Cooper. 1997. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* **17:** 2143–2150.

Du, K., Y. Peng, L.E. Greenbaum, B.A. Haber, and R. Taub. 1997. HRS/SRp40-mediated inclusion of the fibronectin EIIIB exon, a possible cause of increased EIIIB expression in proliferating liver. *Mol. Cell. Biol.* **17:** 4096–4104.

Eperon, I.C., D.C. Ireland, R.A. Smith, A. Mayeda, and A.R. Krainer. 1993. Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *EMBO J.* **12:** 3607–3617.

Fu, X.-D. 1993. Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* **365:** 82–85.

Fu, X.-D. and T. Maniatis. 1992. The 35-kD mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3′ splice site. *Proc. Natl. Acad. Sci.* **89:** 1725–1729.

Fu, X.-D., A. Mayeda, T. Maniatis, and A.R. Krainer. 1992. General splicing factors SF2 and SC35 have equivalent activities in vitro and both affect alternative 5′ and 3′ splice site selection. *Proc. Natl. Acad. Sci.* **89:** 11224–11228.

Ge, H. and J.L. Manley. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* **62:** 25–34.

Ge, H., P. Zuo, and J.L. Manley. 1991. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* **66:** 373–382.

Gontarek, R.R. and D. Derse. 1996. Interactions among SR proteins, an exonic splicing enhancer, and a lentivirus Rev protein regulate alternative splicing. *Mol. Cell. Biol.* **16:** 2325–2331.

Hanamura, A., J.F. Cáceres, A. Mayeda, B.R. Franza, Jr., and A.R. Krainer. 1998. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4:** 430–444.

Heinrichs, V. and B.S. Baker. 1995. The *Drosophila* SR protein RBP1 contributes to the regulation of *doublesex* alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J.* **14:** 3987–4000.

Horton, R.M., H.D. Hunt, S.N. Ho, J.K. Pullen, and L.R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene* **77:** 61–68.

Humphrey, M.B., J. Bryan, T.A. Cooper, and S.M. Berget. 1995. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5′ splice sites in a differential internal exon. *Mol. Cell. Biol.* **15:** 3979–3988.

Irvine, D., C. Tuerk, and L. Gold. 1991. SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* **222:** 739–761.

Jumaa, H., J.L. Guenet, and P.J. Nielsen. 1997. Regulated expression and RNA processing of transcripts from the SRp20 splicing factor gene during the cell cycle. *Mol. Cell. Biol.* **17:** 3116–3124.

Kohtz, J.D., S.F. Jamison, C.L. Will, P. Zuo, R. Lührmann, M.A. Garcia-Blanco, and J.L. Manley. 1994. Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* **368:** 119–124.

Krainer, A.R., G.C. Conway, and D. Kozak. 1990a. The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell* **62:** 35–42.

Krainer, A.R., G.C. Conway, and D. Kozak. 1990b. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes & Dev.* **4:** 1158–1171.

Krainer, A.R., A. Mayeda, D. Kozak, and G. Binns. 1991. Functional expression of cloned human splicing factor SF2: Homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* **66:** 383–394.

Lavigueur, A., H. LaBranche, A.R. Kornblihlt, and B. Chabot. 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes & Dev.* **7:** 2405–2417.

Lawrence, C.E., S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald, and J.C. Wootto. 1993. Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. *Science* **262:** 208–214.

Lynch, K.W. and T. Maniatis. 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes & Dev.* **9:** 284–293.

———. 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes & Dev.* **10:** 2089–2101.

Mayeda, A. and A.R. Krainer. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68:** 365–375.

Mayeda, A. and A.R. Krainer. 1998a. Preparation of Hela cell nuclear and cytosolic S100 extracts for *in vitro* splicing. In *Methods in molecular biology, RNA–protein interaction protocols* (ed. S.R. Haynes). Humana Press, Totowa, NJ. (In press.)

———. 1998b. Mammalian in vitro splicing assays. In *Methods in molecular biology, RNA-protein interaction protocols* (ed. S.R. Haynes). Humana Press, Totowa, NJ. (In press.)

Peng, X. and S.M. Mount. 1995. Genetic enhancement of RNA-processing defects by a dominant mutation in B52, the *Drosophila* gene for an SR protein splicing factor. *Mol. Cell. Biol.* **15:** 6273–6282.

Ramchatesingh, J., A.M. Zahler, K.M. Neugebauer, M.B. Roth, and T.A. Cooper. 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol. Cell. Biol.* **15:** 4898–4907.

Ring, H.Z. and J.T. Lis. 1994. The SR protein B52/SRp55 is essential for *Drosophila* development. *Mol. Cell. Biol.* **14:** 7499–7506.

Screaton, G.R., J.F. Cáceres, A. Mayeda, M.V. Bell, M. Plebanski, D.G. Jackson, J.I. Bell, and A.R. Krainer. 1995. Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J.* **14:** 4336–4349.

Shi, H., B.E. Hoffman, and J.T. Lis. 1997. A specific RNA hairpin loop structure binds the recognition motifs of the *Drosophila* SR protein B52. *Mol. Cell. Biol.* **17:** 2649–2657.

Staffa, A. and A. Cochrane. 1995. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol. Cell. Biol.* **15:** 4597–4605.

Staknis, D. and R. Reed. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14:** 7670–7682.

Sun, Q., A. Mayeda, R.K. Hampson, A.R. Krainer, and F.M. Rottman. 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes & Dev.* **7:** 2598–2608.

Tacke, R. and J.L. Manley. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14:** 3540–3551.

Tacke, R., Y. Chen, and J.L. Manley. 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl. Acad. Sci.* **94:** 1148–1153.

Tanaka, K., A. Watakabe, and Y. Shimura. 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol. Cell. Biol.* **14:** 1347–1354.

Tian, H. and R. Kole. 1995. Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15:** 6291–6298.

Tian, M. and T. Maniatis. 1993. A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell* **74:** 105–114.

———. 1994. A splicing enhancer exhibits both constitutive and regulated activities. *Genes & Dev.* **8:** 1703–1712.

Tuerk, C. and L. Gold. 1990. Systematic evolution of ligands by

exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505–510.

Wang, J. and J.L. Manley. 1995. Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing *in vivo* in diverse ways. *RNA* **1:** 335–346.

Wang, J., Y. Takagaki, and J.L. Manley. 1996. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & Dev.* **10:** 2588–2599.

Watakabe, A., K. Tanaka, and Y. Shimura. 1993. The role of exon sequences in splice site selection. *Genes & Dev.* **7:** 407–418.

Wu, J.Y. and T. Maniatis. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75:** 1061–1070.

Xiao, S.H. and J.L. Manley. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & Dev.* **11:** 334–344.

Xu, R., J. Teng, and T.A. Cooper. 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol. Cell. Biol.* **13:** 3660–3674.

Zahler, A.M. and M.B. Roth. 1995. Distinct functions of SR proteins in recruitment of VI small nuclear ribonucleoprotein to alternative 5′ splice sites. *Proc. Natl. Acad. Sci.* **92:** 2642–2646.

Zahler, A.M., W.S. Lane, J.A. Stolk, and M.B. Roth. 1992. SR proteins: A conserved family of pre-mRNA splicing factors. *Genes & Dev.* **6:** 837–847.

Zahler, A.M., K.M. Neugebauer, W.S. Lane, and M.B. Roth. 1993a. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260:** 219–222.

Zahler, A.M., K.M. Neugebauer, J.A. Stolk, and M.B. Roth. 1993b. Human SR proteins and isolation of a cDNA encoding SRp75. *Mol. Cell. Biol.* **13:** 4023–4028.

Zhang, W.J. and J.Y. Wu. 1996. Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol. Cell. Biol.* **16:** 5400–5408.

Zheng, Z.-M., P. He, and C.C. Baker. 1996. Selection of the bovine papillomavirus type 1 nucleotide 3225 3′ splice site is regulated through an exonic splicing enhancer and its juxtaposed exonic splicing suppressor. *J. Virol.* **70:** 4691–4699.

AT-AC INTRON SPLICING REQUIRES SR PROTEINS.

Michelle L. Hastings and Adrian R. Krainer. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

AT-AC intron excision occurs by a two-step splicing pathway similar to that of conventional GT-AG introns. Despite this similarity, most of the snRNA components of the AT-AC spliceosome are distinct from those of the more common GT-AG spliceosome. This difference likely reflects a need for altered specificity to preserve key RNA:RNA interactions at the splice sites. Whether protein components of the AT-AC spliceosome are novel or shared between the two spliceosomes is not known. SR proteins play important roles in several aspects of splicing of GT-AG introns in higher eukaryotes. To assess a potential role for SR proteins in AT-AC splicing, an assay was developed using HeLa cell S100 extract. Conventional introns are spliced in S100 extract only if SR proteins are also added. Splicing of the SCN4A pre-mRNA AT-AC intron 2 as well as the E2F4 AT-AC intron 3 was not detected in this assay unless an additional nuclear extract fraction was included. In the presence of this fraction and the S100 extract, a dependence on total HeLa SR proteins was observed. AT-AC splicing also was observed using individual recombinant SR proteins. This result demonstrates that conventional SR proteins are required for excision of AT-AC introns. Different SR proteins could promote the splicing reaction, suggesting that SR proteins are functionally redundant in AT-AC splicing. AT-AC intron splicing is stimulated over basal splicing by exon-definition interactions with a downstream conventional 5' splice site or by an exonic splicing enhancer. S100 complementation assays indicate that SR proteins stimulate enhancer- and downstream 5' splice site-dependent splicing as well as basal AT-AC splicing. Whether different SR proteins are required for basal splicing and enhancer or downstream 5' splice site-dependent splicing is being investigated.

SR proteins are required for splicing of AT-AC introns.

Michelle L. Hastings and Adrian R. Krainer. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

AT-AC intron excision occurs by the same two-step splicing pathway as conventional introns. Despite this similarity most of the RNA components of the AT-AC spliceosome are distinct from that of the more common GT-AG spliceosome. This difference likely reflects a need for altered specificity to preserve key RNA:RNA interactions at the splice sites. Whether protein components of the AT-AC spliceosome are novel or shared between the two spliceosomes is not known. To assess a role for SR proteins in AT-AC splicing, an assay was developed using HeLa S100 extract. Splicing of the SCN4A AT-AC intron in S100 is dependent on the presence of total SR proteins and an additional nuclear extract fraction. AT-AC splicing was also observed using individual recombinant SR proteins. This result demonstrates that conventional SR proteins are required for excision of AT-AC introns. Additionally, SR protein activity in AT-AC splicing appears to be functionally redundant. AT-AC intron splicing is stimulated over basal splicing by exon-definition interactions with a downstream conventional 5'ss or by a splicing enhancer. S100 complementation assays indicate that SR proteins stimulate enhancer- and downstream 5'ss- dependent splicing as well as basal AT-AC splicing. Whether different SR proteins are required for basal splicing and enhancer or downstream 5'ss- dependent splicing is being investigated. Current efforts to determine the requirements for additional GT-AG spliceosomal factors in AT-AC splicing will also be presented.